

ESSAYS ON FIRM BEHAVIOR AND THE INTERPRETATION OF QUANTILE REGRESSION

by

Ryutah Kato

**A dissertation submitted to The Johns Hopkins University
in conformity with the requirements for the degree of
Doctor of Philosophy**

Baltimore, Maryland

September, 2020

© 2020 by Ryutah Kato

All rights reserved

Abstract

Chapter 1 provides a dynamic model in which the CEO can manipulate their performance measure at the expense of firm value. The model features learning about the CEO's skill, dismissal and short-termism. While the standard principal agent model focus on the conflict of interest between the principal and the agent regarding the agent's effort level, I focus on the agent's effort allocation between short-term and long-term. There are two potential inefficiencies that could be caused by the conflict of interest between the CEO and the firm. First, agent putting more focus on short-term performance can deteriorate the firms future profitability. In addition, boosting short-term performance without being noticed by the firm can affect the firm's belief about the CEO's skill. This could prevent firms from making optimal firing decisions. I find that inefficiency caused from firms not able to update the skill of the CEO is marginal and most of the inefficiency comes from CEO sacrificing long-term profitability for the sake of short-term performance.

Chapter 2 shows that the slope parameter of the linear quantile regression measures a weighted average of the local slopes of the conditional quantile function. Extending this result, we also show that the slope parameter measures a weighted average of the partial effects for a general structural function. Our results support the use of linear quantile regressions for causal inference in the presence of nonlinearity and multivariate unobserved heterogeneity. The same conclusion applies to linear regressions.

Chapter 3 shows that the slope of the linear quantile regression minimizes the weighted mean squared error loss function of the specification error when the variable of interest is endogenous. Extending this result we also show that the slope of the linear quantile regression measures a weighted average of the derivative of the true quantile regression.

Reader and Advisor: Richard Spady

Reader and Advisor: Yingyao Hu

Acknowledgments

I am extremely grateful to my advisor Richard Spady for many detailed conversations and comments on my work and for continued guidance and encouragement. I would have been lost countless times if not for his help. I wish to thank my co-author Yuya Sasaki for showing an excellent example of working together in academia. I am also grateful to Greg Duffee, Yingyao Hu and Yuya Takahashi for their invaluable advice. Finally I must express my gratitude to my parents, Hideaki Kato and Satomi Kato for providing me continuous support throughout my years of study.

Statement of Co-Authored Work

I confirm that chapter 2 was jointly co-authored with Yuya Sasaki and I contributed at least 50 percent of this work.

Table of Contents

Abstract	ii
Acknowledgments	iv
Statement of Co-Authored Work	v
Table of Contents	vi
List of Tables	x
List of Figures	xi
1 Managerial Short-termism and Firm Value	1
1.1 Introduction	1
1.1.1 Relation to the Existing Literature	4
1.2 The Model	5
1.2.1 The Board's problem	9
1.2.1.1 CEO's problem	11
1.3 Simulation	15
1.3.1 Firing policy	16

1.3.2	CEO turnover	17
1.3.3	CEO's Action	20
1.3.4	Comparative Statics	25
1.3.5	The Cost of Short-term Action	27
1.4	Conclusion	30
	Bibliography	32
2	On Using Linear Quantile Regressions	
	for Causal Inference	34
2.1	Introduction	34
2.2	Relationship between Linear and Nonlinear Functions	36
2.2.1	General Setting	36
2.2.2	Intuition	39
2.2.3	Linear QR and CQF	41
2.2.4	Linear Regressions and CEF	42
2.3	Extensions	45
2.3.1	The Result for the Case of Discrete Regressors	45
2.3.2	The Result for the Case of Multivariate Regressors	47
2.4	Linear QR and Structural Functions	49
2.4.1	Monotone Structural Functions	49
2.4.2	General Structural Functions	50
2.5	Linear Regressions and Structural Functions	53

2.6	Conclusion	54
	Bibliography	56
3	IV Quantile Regression under Misspecification	58
3.1	Introduction	58
3.2	Linear Quantile regression and Quantile regression	59
3.2.1	Weighted Mean Squared Error Minimization	60
3.2.2	Weighted Average Interpretation	61
3.2.3	Discrete Case	63
3.3	Conclusion	65
	Bibliography	66
A	Mathematical Appendix of Chapter 1	68
A.1	Proof of Proposition 1	68
B	Mathematical Appendix of Chapter 2	71
B.1	Proof of Lemma 1	71
B.2	Proof of Theorem 1	75
B.3	Proof of Lemma 2	76
B.4	Proof of Theorem 3	79
B.5	Proof of Lemma 3	80
B.6	Proof of Lemma 4	83
B.7	Proof of Theorem 4	84

B.8	Proof of Lemma 6	85
C	Mathematical Appendix of Chapter 3	87
C.1	Proof of Theorem 7	87
C.2	Proof of Lemma 7	88
C.3	Proof of Theorem 8	90
C.4	Proof of Lemma 8	91
C.5	Proof of Theorem 9	93

List of Tables

1.1	Effects of Short-term Actions on Firm Value	29
-----	---	----

List of Figures

1.1	Firing Threshold for each $C_f = 8, 10$	16
1.2	Firing Threshold for each $\phi = 0.8, 0.95$	17
1.3	CEO Survival rate when $a_t = 0$ for all t for each $C_f = 8, 10$	18
1.4	CEO Survival rate when $a_t = 0$ for all t for each $\phi = 0.8, 0.95$	19
1.5	Unconditional firing percentage of CEO when $a_t = 0$ for all t for each $\theta = 0, 1, 2$	19
1.6	CEO actions when $\eta_t = 0$ for all t for $\theta = -1, 0, 1, 2, 2.5$	21
1.7	Average actions taken by the CEO. The average is taken among 10000 $\{\eta_t\}_{t=1}^{60}$ paths	22
1.8	CEO Survival rate of the Model case and the Benchmark case	23
1.9	Unconditional firing percentage of CEO with skill $\theta = 0, 1, 2$ when the CEO is allowed to engage in short-term action.	24
1.10	Average actions taken by the CEO for each $C_a = 3, 4$. Average is taken among θ and $\{\eta_t\}_{t=1}^{60}$	25
1.11	Average actions taken by the CEO for each $C_f = 8, 10$. Average is taken among θ and $\{\eta_t\}_{t=1}^{60}$	26
1.12	Average actions taken by the CEO for each $\phi = 0.8, 0.95$. Average is taken among θ and $\{\eta_t\}_{t=1}^{60}$	26

1.13	Value of the firm	27
2.1	Illustrations of why M and L must have at least two intersection points. Panel (a) shows a case where there is no intersection point. Panels (b) and (c) show cases where there is only one intersection point. In each of these cases, moving L to \tilde{L} may reduce the weighted mean squared distance from M , and hence L cannot minimize the weighted mean squared distance.	40

Chapter 1

Managerial Short-termism and Firm Value

1.1 Introduction

Most modern corporations are hierarchical, with the result that its most important employee, the Chief Executive Officer (CEO), has an major influence on the firm's success. At the same time, since the CEO can appoint the firm's board of directors and is able to control the information that the board receives, the CEO can attenuate supervision from the board. A conscientious board can seek to mitigate these well-known effects by a variety of measures that seek to evaluate the CEO objectively. Since corporations are primarily profit-seeking enterprises, one of the easiest and most relevant measures of CEO performance and competence is firm profitability. However, there are many methods of inflating short-term profits and strict performance standards may have the unintended consequence of inducing a CEO to engage in short-term behavior that enhances immediate profit at the expense of the long-term profitability that serves the interests of the firm's owners. Ideally, the board is able to perfectly observe the CEO's actions to correctly evaluate the CEO. However, when this ideal is not met, the board's evaluation about the CEO can be easily distorted by the CEO's short-term behavior that enhances immediate profits. This paper examines the amount of shareholder

value at stake from the CEO's behavior.

The removal of poorly performing CEO is essential to maximize firms performance. The firm should fire the CEO whenever it is profitable to do so. However the threat of being replaced may induce the CEO to engage in short-term behavior that enhances immediate profit at the expense of the long-term profitability. This could prevent the board from making the optimal replacement decisions. This paper provides a dynamic model in which the CEOs can manipulate their performance measures at the expense of firm value. In the model, the board decides at each period whether to fire or keep the current CEO. CEOs have different skill levels, meaning that some CEOs are more skilled than others. The CEO with a higher skill can generate higher average profitability. The board cannot directly observe CEO's skill, but instead learns about it over time by observing a noisy signal of the skill, profits. Each period the board observes the signal and updates its belief about the CEO's skill and replaces the current CEO with a new CEO with unknown ability whenever it is optimal to do so. However, the CEO can manipulate this signal by engaging in a short-term action. Short-term action enhances the current period's profit at the expense of long-term profitability. Short-term action involves actions such as cutting R&D and reducing employee training. Also, it could be interpreted as spending resources on earning management. By enhancing the current period's profit, the CEO is able to make the board believe that his skill is higher than it actually is. When the CEO is in danger of being replaced, the incentive to engage in short-term action to increase the chance of surviving this period increases even if it means it would deter the firm value. Complicating matters, the board cannot observe the amount of short-term action taken by the CEO. Hence, the short-term action not only deters firm value but also makes it more difficult for the board to correctly update the belief about the CEO's skill. Thus the board cannot make an optimal firing decision. Conflict of interest between the board and the CEO arises since the board only cares about the value of the firm, but the CEO

cares both about his job security and the value of the firm. This implies that the firm and the CEO having conflicting goals which potentially leads to inefficiency. The goal of this paper is to analyze the CEO's behavior and how the conflict of interest between the board and the CEO affect the value of the firm.

This model can be interpreted as a dynamic principal agent model with learning. The standard principal agent model often assumes that the principal knows the correct distribution of future output and focus on the optimal contract that induces the agent to take the desired level of effort by the principal. However, in this case the agent can mislead the principal about the future by choosing the action that the principal does not expect. Demarzo and Sannikov (2016) analyze a dynamic principal agent model which analyzes the case where the principal and the agent both learns about the fundamentals of the firm. Their model focus on the conflicting goals between principal and the agent regarding the agent's effort level. The agent can distort principals beliefs about the fundamentals of the firm by not exerting the effort level that the principal desires. They take into account the consequences of this information rent and characterize the dynamics of the agent's compensation, firm payout and liquidation/termination policies. Instead of focusing on the conflict of interest between principal and the agent regarding the agent's effort level, I focus on the conflict of interest regarding the allocation of the agent's effort between short-term and long-term. Since a very high work ethic is required to become a CEO in the first place, it is natural to think that their cost of effort is very low. If the cost of effort of the agent is negligible then there will not be a major conflict of interest regarding the level of effort. However, there could be a conflict between the CEO and the firm regarding how the CEO should allocate his effort. The CEO may allocate more effort on projects that has immediate reward since that would increase the short-term performance of the firm which could result in an increase in job security.

In section 1.2, I present a model that captures these features and in section 1.3, I show a numerical simulation of the model. Model predicts that the lower skilled CEO is more likely to focus on short-term performance compared to the higher skilled CEO. This is intuitive since the lower skilled CEO is more in danger of getting fired compared to the higher skilled CEO. As a result, the lower skilled CEO focus more on the short-term performance in order to keep his job instead of focusing on increasing the long-term value of the firm. Higher skilled CEO can focus more on increasing the long-term value of the firm since his job is more secure. Also, the model predicts that the CEO is more likely to focus on the short-term when they are at risk of losing their job. This is intuitive since if the CEO job is secure, he can focus more on increasing the value of the firm which increases the share values that the CEO owns. In addition, I find that inefficiency caused from firms not able to update the skill of the CEO is marginal and most of the inefficiency comes from CEO sacrificing long-term profitability for the sake of short-term performance.

1.1.1 Relation to the Existing Literature

Although it is difficult to prove that such CEO behavior exists, there are researches done to provide some evidence of managerial short-termism. Graham, Harvey and Rajgopal (2005) conducted a survey and interviewed 400 executives. 80 percent of the survey participants reported that they would decrease discretionary spending on R&D, advertising, and maintenance to meet an earnings target. Also 55 percent of the participants reported that would delay starting a new project to meet an earnings target, even if such a delay entailed a small sacrifice in value.

Edmans, Fang and Lewellen (2015) study the quantity of equity scheduled to vest in a given year. They find that vesting equity is significantly negatively correlated with cuts in R&D, advertising, and capital expenditure, and positively correlated with the likelihood of

meeting or narrowly beating earnings targets.

There could be many reasons for the CEO to focus on the short-term performance at the expense of long-term firm value. For example, CEO may focus more on short-term performance to enhance short-term stock prices(see Stein (1989), Bebchuk and Stole (1993)) or to enhance their reputations in the CEO labor market(see Campbell and Marino (1994).) In this paper, the threat of getting replaced induces the CEO to focus on the short-term performance. Dismissal can lead to loss of current employment, reduced future career options(see Brickley et al. (1999)), loss of unvested equity-based compensation (see Dahiya and Yermack (2008), Laux (2012)) and so on. It is natural to expect CEOs to do everything to avoid getting fired.

The model presented in this paper features moral hazard problem with learning. While the standard principal agent model often assumes that the principal correctly knows the correct distribution of future output, I consider a model where the agent can mislead the principal about the future output by choosing the action that the principal does not expect. Demarzo and Sannikov (2016) analyze a model where the principal and the agent both learn about the fundamentals of the firm. While their model focus on the conflicting goals between principal and the agent regarding the agent's effort level, I focus on the conflict of interest regarding the allocation of the agent's effort between short-term and long-term instead. Taylor (2010) uses a similar model that features learning about CEO ability and costly turnover to estimate the cost of firing that would rationalize observed turnover rates. However, his model does not include any hidden action by the CEO.

1.2 The Model

In this section I study a dynamic model where the CEO can boost the firm's short-term performance at the expense of long-term value of the firm. The model features firm which

lives infinitely, the CEO, and the board. In the model, each CEO has different skill levels θ . Some CEOs are more skilled than others, meaning they can generate higher average profitability. The board cannot directly observe CEOs skill, but learns gradually by observing a noisy signal, which is the profitability. The board infers the CEO's skill by observing the profitability and makes an optimal firing decision. However, the CEO can engage in a short-term action a_t to boost the current profitability. This short-term action is only known by the CEO and not observed by the board. Profit π_t is given as follows

$$\pi_t = h_t + C_a \cdot a$$

$$h_t = \phi \cdot h_{t-1} + \theta - a_t + \eta_t$$

Profit π_t consists of two components h_t and $C_a \cdot a$. h_t denotes the quality of the firm at time t and $C_a \cdot a_t$ is the profit generated by the CEOs short-term action $a_t \in [0, A]$, $A > 0$. If the CEO does not engage in short-term action i.e. $a_t = 0$, the quality of the firm would be the same as the observed profit π_t . Notice that the short-term action would increase the current profit by $(C_a - 1) \cdot a_t$ but decreases the quality of the firm by a_t . Short-term action can be interpreted as an action that has negative effect on the long-term value of the firm but results in immediate increase in current profit. These actions may involve cutting R&D and reducing employee training. Also, short-term action a_t could be interpreted as allocating resources to earnings manipulation instead of allocating effort to projects that actually increase firm value. Parameter ϕ reflects the persistence in firm quality h_t with $0 < \phi < 1$. This persistence allows the short-term action to have a long-term consequence. The shock η_t is independently and normally distributed with mean zero and variance σ_η . θ denotes the skill of the CEO which is constant over time. The skill of the CEO θ is a private information and only known to the CEO himself. The board represents the interests of shareholders and is replacing the CEO if

necessary. Each period the board makes a firing decision $d_t \in 0, 1$ to maximize the firm value. Whenever the firm fires a CEO, they incur a firing cost C_f and replace him with a new CEO. The cost of firing includes executive search costs, severance packages, adjustment costs and so on. They maximize the firm value M_t which is given by

$$M_t = E_t \left[\sum_{s=0}^{\infty} \beta^s \cdot \pi_{t+s} - d_{t+s} \cdot C_f \right] \quad (1.2.1)$$

β is the discount factor with $0 < \beta < 1$. Now I make an additional assumption on the parameter C_a .

Assumption 1. (i) $1 < C_a < \frac{1}{1 - \beta \cdot \phi}$ for all $a \in \mathbb{R}_+$

Since engaging in short-term action moves the current profit by $(C_a - 1) \cdot a_t$, assumption 1 ensures that positive short-term action $a_t > 0$ will increase current period's profit. Also short-term action decreases the health of the firm by a_t which decreases the firm's value by $\frac{a_t}{1 - \beta \cdot \phi}$. Assumption 1 ensures that $\frac{a_t}{1 - \beta \cdot \phi} > C_a \cdot a_t$ which implies that any positive short-term action $a_t > 0$ is harmful for the firm value. Therefore $a_t = 0$ will be the optimal for the firm's value.

The board can observe all parameters, but cannot observe CEO's skill levels and CEO's action. Therefore, when the board observes high profit, it cannot tell whether this is due to CEO skill, luck or the short-term action. When the board hires a new CEO, the prior belief of the CEO's skill is given as

$$\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

Each period the board updates its belief about skill θ based on the information contained in π_t .

The CEO plans to work at a firm for a finite time period, $t \in 0, 1, \dots, T$ where T denotes the retirement date. I assume that the retirement date T is known by both the CEO and the board.

The CEO chooses a_t from $[0, A]$ to maximize following.

$$textE_t[\sum_{s=0}^{T-t} \beta^s \cdot (d_{t+s} - \alpha \cdot \phi^{T-s+t} \cdot a_{t+s})] \quad (1.2.2)$$

$d_t \in \{0, 1\}$ denotes whether the CEO is working or not. $d_t = 1$ if the CEO is working and $d_t = 0$ if the CEO is not.

The objective of the CEO consists of 2 components. First component is the expected duration that the CEO can stay on the job. The second component is the disutility that the CEO incur from the short-term action. This objective function can be interpreted as follows. The CEO acquire benefits from working as a CEO, hence the CEO gains utility from keeping his job every period. Also, I assume that the CEO owns some shares of the firm. Engaging in short-term action harms the long-term profitability of the firm, hence decreases the value of the shares that the CEO earns. Therefore, the CEO incur some disutility from engaging in short-term action. For simplification, I assume that the CEO cares about the firm's value at time T . This could be interpreted as he is granted a stock option that vests after he retires. However, engaging in short-term action could increase the probability of keeping the job. Hence, there is a some trade-off between job security and the value of the shares. This utility function of the CEO implies that the CEO is only focused on two things. First is the expected discounted duration that he can stay at the job. Since the CEO does not only receive high wages and bonuses but is also granted a special treatment and privileges. It is natural for the CEO to maximize the duration that he can stay on the job. The CEO is also worried about the value of the firm since he owns some equity which is tied to the firm's value. By assuming the CEO's utility function this way, I am treating the compensation package that the CEO receives is determined exogenously. While in the standard principal agent model, principal can chose a contract to offer to the agent to mitigate moral hazard, I assume that compensation package is decided from the exogenous factor such as competition from other firms to sign the CEO.

This assumption allows me to simplify the situation and focus on how the board's termination policy affects CEO's behavior and the value of the firm.

For simplicity I also make the following assumption.

Assumption 2. *If there exist $s > 0$ such that $d_s = 0$ then $d_t = 0$ for all $t > s$.*

Assumption 2 implies that once the CEO loses his job, he retires and never works again. At first glance this assumption might seem too extreme. However, it is natural to think that getting fired from the job leads to a massive decline in reputation. As a result it will be extremely difficult for the fired CEO to find a new job that gives a similar level of utility to the previous job. Hence, Assumption 2 implies that once the CEO is fired, he can only find a job that is much worse than his previous that he would rather retire.

1.2.1 The Board's problem

For simplicity, I assume a naive board that believes the CEO always takes the action that maximizes the firm value i.e. $a_t = 0$. This assumption means that the short-term action taken by the CEO is not just unobserved but also unexpected to the board. Although this assumption may seem a little extreme at first, there are plenty of real life examples where the board completely failed to detect CEO's myopic behavior. Under this assumption, the board believes that the profit process is given as

$$\pi_t = \phi \cdot \pi_{t-1} + \theta + \eta_t$$

Based on the π_t the board updates its belief about the CEOs' skill θ according to Bayes' rule. Since prior beliefs and signals are normally distributed, the boards' posterior beliefs about CEOs' skill will also be normally distributed. At the end of period t , distribution of the

board beliefs are denoted as

$$\theta \sim \mathcal{N}(\mu_t, \sigma_t^2)$$

Applying Bayes rule the posterior variance follows

$$\sigma_t^2 = \sigma_0^2 \cdot \left(1 + t \cdot \frac{\sigma_0^2}{\sigma_\eta^2} \right)^{-1} \quad (1.2.3)$$

which goes to zero in the limit when t becomes infinite. The posterior mean follows

$$\mu_{t+1} = \mu_t + \frac{\sigma_t^2}{\sigma_\eta^2} \cdot (\pi_{t+1} - \phi \cdot \pi_t - \mu_t) \quad (1.2.4)$$

The posterior mean can be written as the sum of the mean belief of the previous period and the current periods unexpected profit.

Objective of the board is to maximize the firm value defined in equation(1.2.1) At the beginning of each period the firm decides to keep or fire the CEO. The proposition below provides the Bellman equation for the board's optimization problem. Let τ_t denote the number of years the current CEO worked for the firm at the end of time t .

Proposition 1. *The firm value at the beginning of period $t + 1$, M_{t+1} can be written as*

$$M_{t+1} = \frac{\phi}{1 - \beta \cdot \phi} \cdot \pi_t + V(\mu_t, \tau_t) \quad (1.2.5)$$

where the value function V solves the Bellman equation

$$V(\mu, \tau) = \max\{V_f, V_k\}$$

$$V_f = V(\mu_0, 0) - C_f$$

$$V_k(\mu, \tau) = \begin{cases} \frac{1}{1 - \beta \cdot \phi} \mu + \beta \cdot E[V(\mu + \frac{\sigma^2(\tau)}{\sigma_\eta^2} \cdot \epsilon(\tau), \tau + 1)] & \text{if } 0 \leq \tau \leq T - 1 \\ V(\mu_0, 0) & \text{if } \tau = T \end{cases}$$

$$\epsilon(\tau) \sim \mathcal{N}(0, \sigma_\eta^2 + \sigma^2(\tau))$$

The proof of this proposition is given in the Appendix. Extensive part of the proof follows Taylor(2010). Proposition 1 implies that there exists a threshold $\mu^*(\tau)$ such that the board fires the CEO as soon as its belief of the CEO's skill i.e. the posterior mean of θ , μ_t , drops below a threshold $\mu^*(\tau)$. This threshold depends on all the parameters and the number of years the current CEO has worked for the firm. For example, raising the cost of firing will reduce the threshold, making firing less likely.

1.2.1.1 CEO's problem

Each period CEO chooses $a_t \in [0, A]$ to maximize the objective function give in equation (2.5) knowing the firing policy of the board. Let $\mu^*(\tau_t)$ denote the firing threshold at time t and the CEO has worked for τ_t periods. Then the CEO is fired at period t is if

$$\mu_t < \mu^*(\tau_t) \Leftrightarrow \eta_t < \frac{\sigma_\eta^2}{\sigma_{t-1}^2} \cdot (\mu^*(\tau_t) - \mu_{t-1}) - (\theta + C_a \cdot a_t - a_t - \phi \cdot C_a \cdot a_{t-1})$$

Let p_t denote the probability of getting fired at time t . Then the probability p_t is given as

$$p_t = F_\eta \left(\frac{\sigma_\eta^2}{\sigma_{t-1}^2} \cdot (\mu^*(\tau_t) - \mu_{t-1}) - (\theta + C_a \cdot a_t - a_t - \phi \cdot C_a \cdot a_{t-1}) \right)$$

where F_η is the cdf of η . Note that the probability of getting fired p_t increases in σ_{t-1} , a_{t-1} and decreases in θ , a_t . This is intuitive since a higher variance σ_{t-1} implies that the signal will have a higher impact towards the mean belief about the CEOs skill, which means that the bad luck will decrease μ_t heavily. Also the probability of getting fired would increase in a_{t-1} . Since the board assumes that $a_{t-1} = 0$, profit function that the board expects at period t is $\phi \cdot \pi_{t-1} + \theta + \eta_t$. However, the actual profit function is $\phi \cdot \pi_{t-1} + \theta + \eta_t - \phi \cdot C_a \cdot a_{t-1}$ which is $\phi \cdot C_a \cdot a_{t-1}$ lower than the board expects. Therefore it is harder for the CEO to meet the expectation of the board.

The CEOs problem can be written as

$$\max \mathbb{E}_t \left[\sum_{s=0}^{T-t} \beta^s \cdot (1 - \alpha \cdot \phi^{T-s+t} \cdot a_{s+t}) \cdot \prod_{j=t}^{t+s} (1 - p_j) \right]$$

s.t.

$$\mu_{t+1} = \mu_t + \frac{\sigma_t^2}{\sigma_\eta^2} \cdot (\pi_{t+1} - \phi \cdot \pi_t - \mu_t)$$

$$p_t = F_\eta \left(\frac{\sigma_\eta^2}{\sigma_{t-1}^2} \cdot (\mu^*(\tau_t) - \mu_{t-1}) - (\theta + C_a \cdot a_t - a_t - \phi \cdot C_a \cdot a_{t-1}) \right)$$

Value function satisfies

$$V(\theta, \mu_{t-1}, a_{t-1}) = \max_{a_t} \{1 - \alpha \cdot \phi^{T-t} \cdot a_t + (1 - p_t) \cdot E_t[V(\theta, \mu_t, a_t)]\}$$

$$V(\theta, \mu_T, a_T) = 0$$

This is solved by backwards induction. As mentioned previously, the CEO faces a tradeoff of job security and firm value. Engaging in short-term action would decrease the value of the shares that the CEO owns but increases the job security this period.

This is the source of the conflict of interest between the board and the CEO. While the board only cares about the value of the firm, the CEO cares both about his job security and the value of the firm. This implies that the firm and the CEO having conflicting goals which potentially leads to inefficiency.

The board wants to replace the CEO with low skill to maximize firm value which is given by (1.2.1). Since the skill of the CEO is not observed by the board, the board has to infer about the CEO's skill from the realized profits. However, the threat of being replaced may induce the CEO to focus more on the short-term performance and sacrifice long-term profitability to manipulate the board's belief about the CEO's skill. Since the board does not observe the CEO's action a_t and updates the CEO's skill based on the assumption that $a_t = 0$, the CEO may have an incentive to engage in short-term action to boost the short-term profit in order to make the board believe that he is high skilled. For example, when the board's beliefs about the CEO's skill is close to the firing threshold, the CEO may engage in short-term action just to survive this period. There are two inefficiencies that could be caused by the short-term action. First, agent putting more focus on short-term performance can deteriorate the firm's future profitability. When the CEO engages in short-term action $a_t > 0$ to increase his job security, the firm's value will decrease by $(\frac{1}{1-\beta \cdot \phi} - C_a) \cdot a_t$. From Assumption 1, any positive

short-term action will decrease firm's value. In addition, boosting short-term performance without being noticed by the firm can affect the board's beliefs about the CEO's skill. This could prevent firms from making optimal firing decisions. Recall that the board cannot observe CEO's skill and the board uses Baye's rule to update its beliefs about the CEO skill using the realized profits assuming that the CEO does not engage in short-term action. Since the board makes a firing decision based on these beliefs to maximize firm value given by (1.2.1), not able to update the CEO's skill properly can have an effect on firm value. In the next section, in order to understand how the conflict of interest between the board and the CEO affects firm value, I solve the model numerically.

1.3 Simulation

In this section I solve the model numerically and discuss the model prediction about firing policy, frequency of CEO turnover, the CEO action and the inefficiency caused from the CEO's action. To solve the board's problem, I first derive the board's beliefs about CEO skill by using Bayes's rule. Then I substitute this into the board's objective function (1.2.1) and numerically compute the Bellman equation(1.2.5). Next I substitute the firing threshold obtained from numerically solving the board's problem to the CEO's objective function (1.2.2) . I numerically solve(1.2.5) by backward induction to obtain the optimal CEO action. Throughout the simulation I use the parameter values $\beta = 0.99$, $\mu_0 = 1$, $\sigma_0 = 1$, $T = 60$. Also, I regard 1 period as 1 quarter. Hence, the yearly discounted factor would be $0.99^4 \approx 0.96$. Also, $T = 60$, which implies that CEO retires after 15 years in office.

The simulation goes as follows.

1. The new CEO is hired and his skill θ is drawn from $N(0, 1)$
2. The CEO decides the action a_t .
3. η_t is drawn from $N(0, \sigma_\eta)$ and the profit π_t is realized.
4. Given π_t the board updates its beliefs based on (1.2.3) and (1.2.4).
5. If the posterior mean is above the firing threshold the board keeps the CEO and returns to (2)
6. If the posterior mean is below the firing threshold the board fires the CEO and returns to (1)

1.3.1 Firing policy

In this section I illustrate the board's firing policy. The board updates the CEO's skill every period by observing the profit π_t every period. The board fires the CEO whenever the posterior mean drops below the threshold derived from the optimal firing policy. This threshold depends on the model parameters and the CEO tenure.

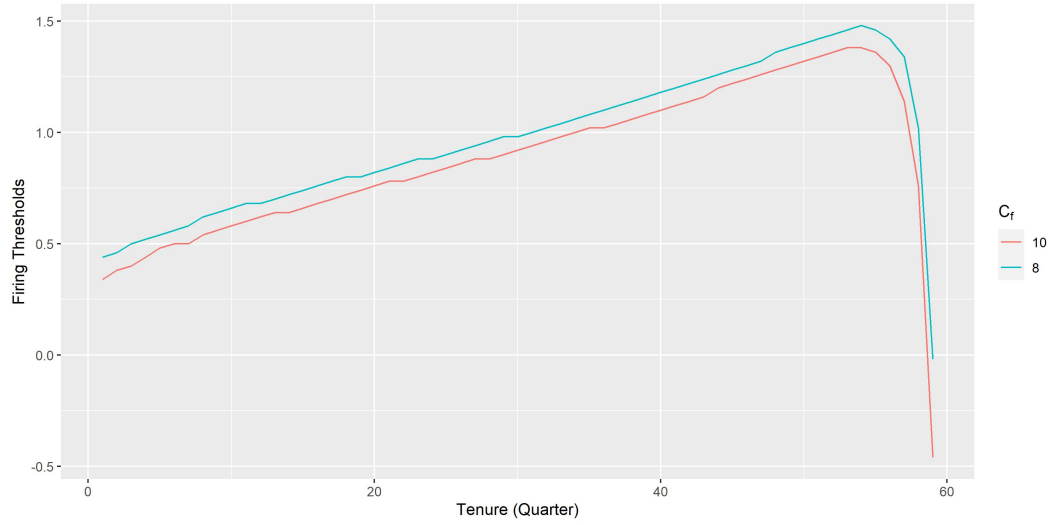


Figure 1.1: Firing Threshold for each $C_f = 8, 10$

Figure 1.1 plots the firing threshold versus CEO tenure for 2 values of firing cost $C_f = 8, 10$ when the persistence parameter $\phi = 0.8$. Notice that firing threshold increases with tenure. This means that the board is more willing to fire CEO's who have worked longer for the firm. There is less uncertainty about the long tenured CEO. The board prefers higher uncertainty about the CEO's skill since they can minimize the downward risk by firing him. The option to fire a CEO is more valuable when the uncertainty about the CEO's skill is higher. Also the firing threshold starts to decline towards the CEO's retirement period 60. This is because the board knows that the CEO is going to retire soon anyway. Hence the belief about the CEO's skill has to be lower to make the firing worth it for the board. Increasing the cost of firing will

shift the threshold down and make firing less likely. If firing a CEO is more costly, the belief about the CEO skill has to be lower to make the firing worth it for the board.

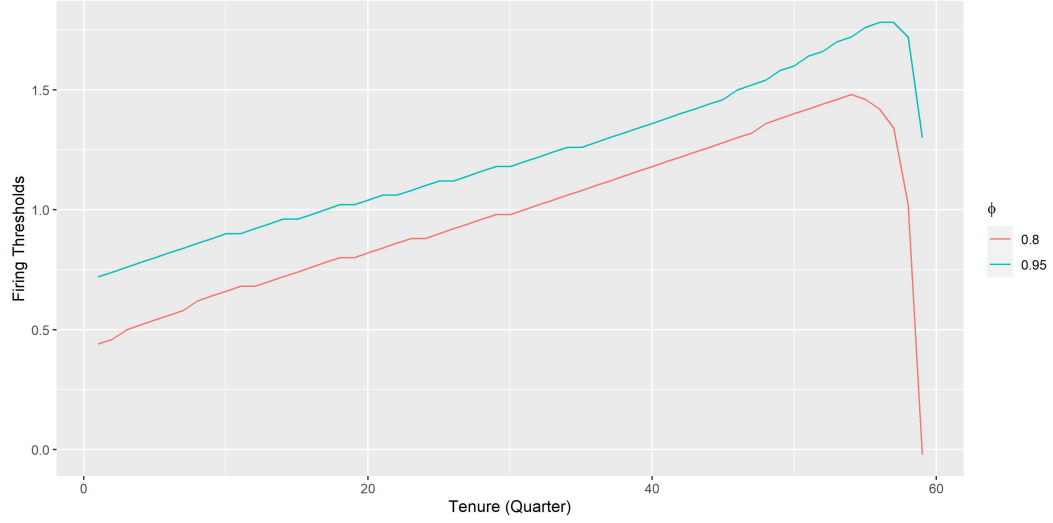


Figure 1.2: Firing Threshold for each $\phi = 0.8, 0.95$

Figure 1.2 shows the firing threshold versus CEO tenure for 2 values of persistence $\phi = 0.8, 0.95$ when the firing cost is $C_f = 8$. Increasing the persistence parameter ϕ will shift the threshold up and make the firing more likely. This is because if the persistence parameter is high, having a low skilled CEO is more punishing to the firm. From equation (1.2.1) the CEO adds θ to the health of the firm every period. Since the health of the firm is persistent adding θ to the health of the firm results in $\frac{\theta}{1-\beta\cdot\phi}$ more firm value. Hence, CEO's skill is more important to firm value when the persistence parameter is high. Higher persistence parameter ϕ leads to more firing since keeping the low skilled CEO will cause more harm to the firm value.

1.3.2 CEO turnover

In order to understand the model better, I illustrate the model prediction on frequency and the timing of firing when the CEO always take the optimal action for the firm i.e. $a_t = 0$ for

all t . I will call this the benchmark case. In this case the board can correctly update the beliefs about the CEO's skill thus can make the optimal firing decision. I use the same parameters in the previous section which are $\beta = 0.99, \mu_0 = 1, \sigma_0 = 1$.

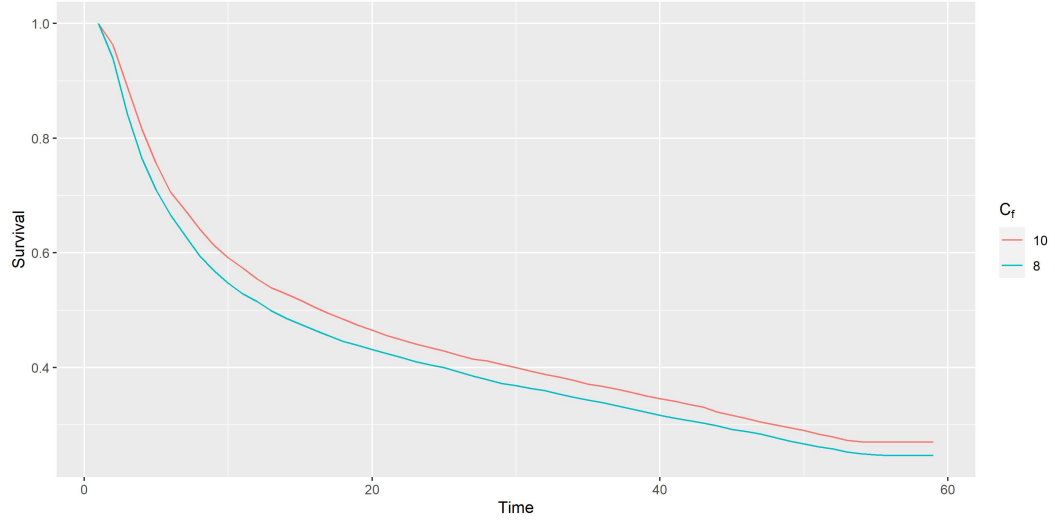


Figure 1.3: CEO Survival rate when $a_t = 0$ for all t for each $C_f = 8, 10$

Figure 1.3 plots the predicted CEO survival rate versus CEO tenure for two values of cost of firing $C_f = 8, 10$. Other parameters are $\beta = 0.99, \mu_0 = 1, \sigma_0 = 1, \sigma_\eta = 3, \phi = 0.8$. As shown in the previous section, increasing the cost of firing will shift the threshold down and make firing less likely. 27 percent of the CEO will survive 60 periods and retire when $C_f = 10$ while only 24 percent can survive for 60 periods and retire when $C_f = 8$. Also increase in the firing cost affects the timing of the firing as well. Firing occurs later when the firing cost is higher.

Figure 1.4 plots the predicted CEO survival rate versus CEO tenure for two values of persistence parameter $\phi = 0.8, 0.95$. The other parameters are $\beta = 0.99, \mu_0 = 1, \sigma_0 = 1, \sigma_\eta = 3, C_f = 8$. Since increasing the persistence parameter $\phi = 0.8, 0.95$ will increase the firing threshold, it makes the CEO firing more likely. 24 percent of the CEO will survive 60 periods and retire when $\phi = 0.8$ while only 14 percent can survive for 60 periods and retire when

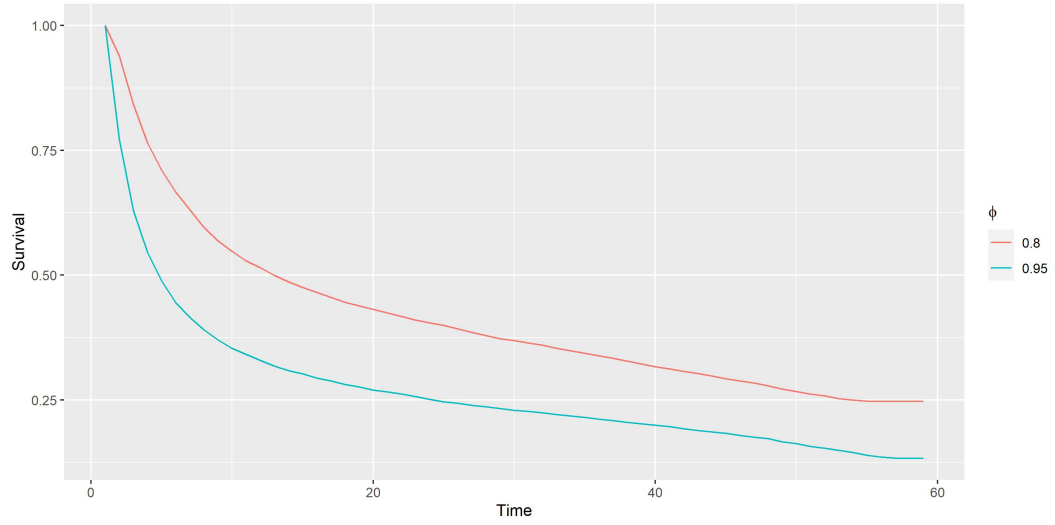


Figure 1.4: CEO Survival rate when $a_t = 0$ for all t for each $\phi = 0.8, 0.95$

$\phi = 0.95$.

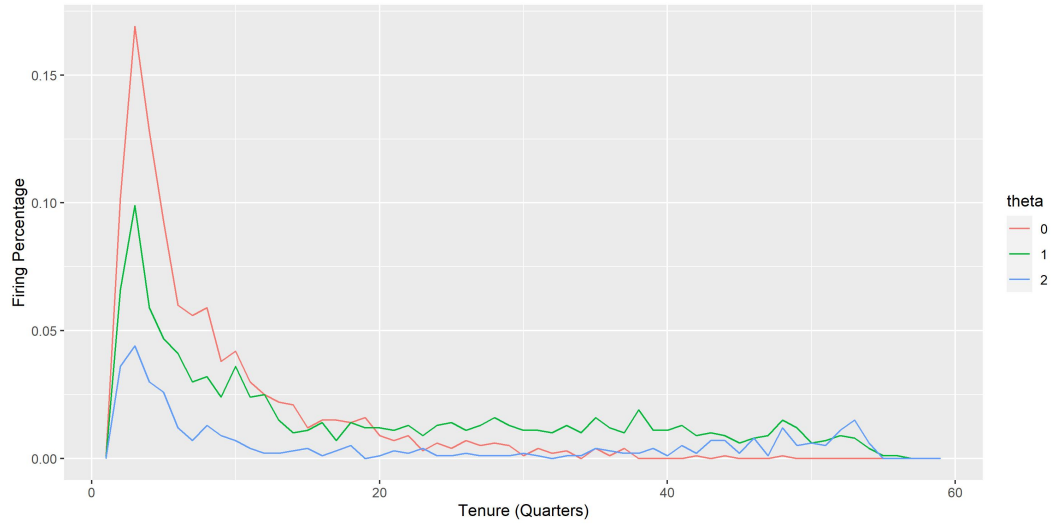


Figure 1.5: Unconditional firing percentage of CEO when $a_t = 0$ for all t for each $\theta = 0, 1, 2$

Figure 1.5 plots the unconditional firing percentage of CEO with 3 different values of skill level $\theta = 0, 1, 2$ at each tenure levels when $a_t = 0$ for all t . Obviously the lower skilled CEO is more likely to get fired compared to the higher skilled CEO. Notice that most of the firing of the CEO occurs in the early stages of their tenure. This is intuitive since the uncertainty about

the CEO's skill is high in the early stage of the CEO's tenure. Hence one bad outcome will have a high impact on the boards beliefs about the CEO's skill which could lead to dismissal. However if the CEO is highly skilled, the board is likely to learn that the CEO is highly skilled over time, thus makes the firing at a later stage of the tenure unlikely.

This could be also seen from equation (1.2.3) and (1.2.4). As shown in equation (1.2.3) ,the variance about the CEO's skill decreases over time as the board learns about the CEO's skill. Equation (1.2.4) shows that as variance σ_t decreases, change in the mean beliefs about the skill decreases as well.

There will still be some firing in the later stage of CEO tenure since firing threshold increases in CEO tenure as shown in the previous section. Even though keeping their job for long period of time usually implies that the board believes that the CEO is highly skilled, the firing threshold also increases overtime which makes it tougher for the CEO to keep his job.

1.3.3 CEO's Action

Now I analyze the model prediction when the CEO is allowed to engage in short-term action. In this case CEO can distort the board's belief by engaging in short-term action. In this section I illustrate the model predictions about the CEO's optimal actions when the parameters are $\beta = 0.99$, $\mu_0 = 1$, $\sigma_0 = 1$, $\phi = 0.8$, $C_f = 8$, $\sigma_\eta = 3$, $C_a = 4$. First I illustrate how the CEOs with different skill level behave when $\eta_t = 0$ for all t . I consider this case to show how the CEOs with different skill level for one specific path of η_t . Since η_t is drawn from $N(0, \sigma_\eta)$, $\eta_t = 0$ could be considered as case where the CEO is neither lucky nor unlucky.

Figure 1.6 plots the actions taken by the CEO versus the CEO tenure for 5 values of skill level θ when $\eta_t = 0$ for all t . For example the green dotted line shows the actions taken

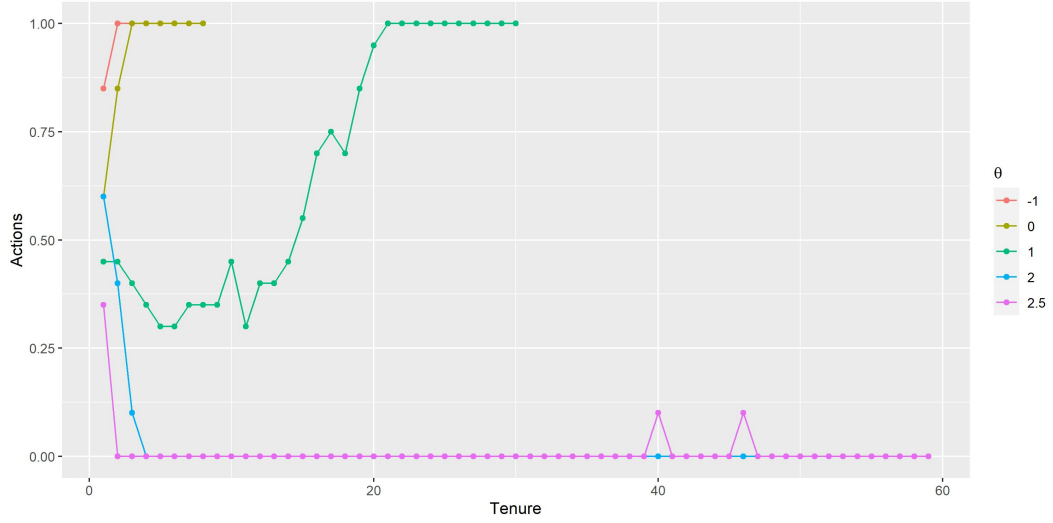


Figure 1.6: CEO actions when $\eta_t = 0$ for all t for $\theta = -1, 0, 1, 2, 2.5$

by skill level $\theta = 1$ CEO when the disturbance in the profits $\eta_t = 0$ for all t . The line ends at period 30 meaning that the CEO is fired after 30 periods. Notice that lower skilled CEO are more likely to engage in more short-term actions compared to the higher skilled CEO on average. This is intuitive since the lower skilled CEO is more in danger of getting fired compared to the higher skilled CEO. Hence it is natural for them to focus on surviving the current period and worry less about the future survivability. On the other hand, higher skilled CEO can focus more on the future survivability and the value of the shares that he owns. This implies that hiring lower skilled CEO may be more costly to the firm than it should be since the lower skilled CEO is more likely to engage in short-term action. Also notice that the CEO plays more short-term action when they are close to getting fired. As the CEO's job become more insecure, the CEO tends to focus more on surviving the current period. For the higher skilled CEO, most of the short-term actions are taken in the early stages of their tenure. This is because they are at most vulnerable in the early stages of their tenure since uncertainty about the CEO's skill from the board's perspective. Hence one bad outcome could swing the boards beliefs about the CEO's skill. Therefore even if the CEO is highly skilled it is natural to protect

himself from getting accidentally fired from bad luck. Once the board learns about the CEO's skill decently well, the outcome does not move the board's belief to that extent. Hence the short-term action will not change the probability of getting fired as much as it will in the early stages of the CEO's tenure. Interestingly, the CEO with $\theta = 2$ engages in more short-term action compared to the CEO with $\theta = 1$ in the first period. This implies that if the CEO is highly skilled, the optimal strategy for him is to engage heavily in short-term action in the early periods to make sure that the board recognize him as a highly skilled CEO. Once that is achieved, he can just focus on the long-term. This strategy may not work for the average skilled CEO since he can only deceive the board for a short period of time and eventually the board will find out that he is an average CEO. Next I illustrate the average actions taken by the CEO among 10000 η_t paths.

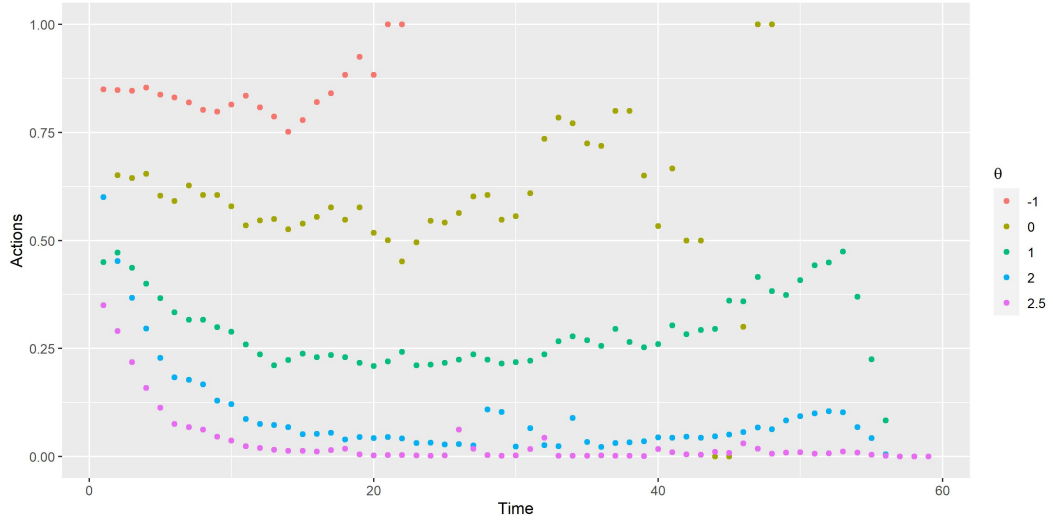


Figure 1.7: Average actions taken by the CEO. The average is taken among 10000 $\{\eta_t\}_{t=1}^{60}$ paths

Figure 1.7 plots the average actions taken by the CEO from 10000 simulations versus the CEO tenure for 5 values of skill level. Similar to the $\eta_t = 0$ for all t case, we can see that on average, the CEO with lower skill tends to engage in more short-term action compared to the CEO with higher skill. Also for the higher skilled CEO, most of the short-term actions

are taken in the early stages of their tenure. Short-term actions starts to increase in the later stages of the CEO's tenure since the firing threshold increases with CEO tenure. As the firing threshold increases over time, the CEO becomes more vulnerable to bad outcomes which leads to increase in short-term actions to protect himself from getting fired.

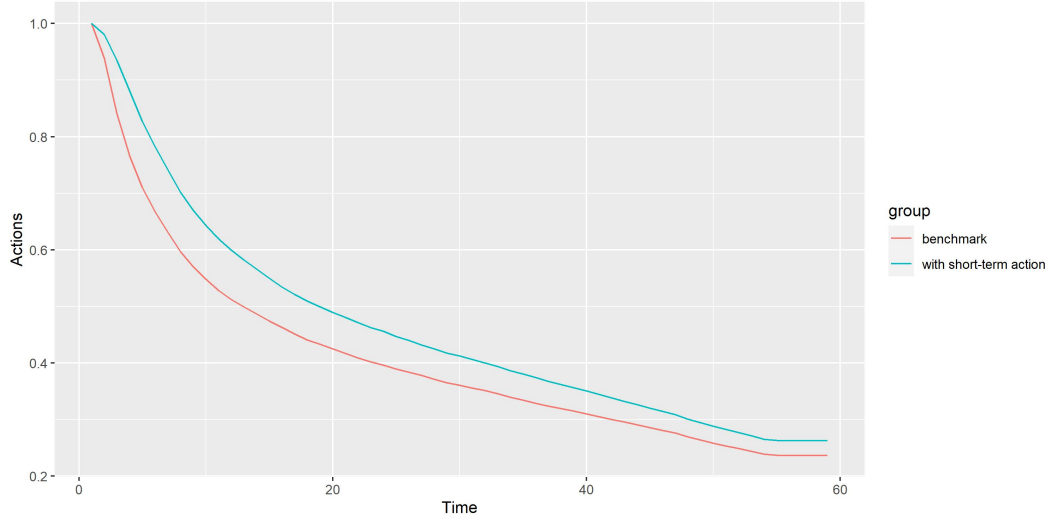


Figure 1.8: CEO Survival rate of the Model case and the Benchmark case

Figure 1.8 plots the survival rate of the CEO versus tenure. Benchmark case which is shown in a red line is the survival rate of the CEO when the short-term action is not allowed. In this case the the board can correctly update the CEO's skill. The blue line shows the survival rate when the CEO can engage in short-term action. Since the CEO can manipulate the boards belief to their advantage, they are able to survive longer.

Figure 1.9 plots the unconditional firing percentage of CEO with 3 different values of skill level $\theta = 0, 1, 2$ at each tenure levels when the CEO is allowed to engage in short-term action. Compared to the bench mark case, there are less firing in the early stages since the CEO has the option to manipulate the outcome. This makes the CEO survive longer in general. Lower skilled CEOs surviving longer than they should will harm the value of the firm. However

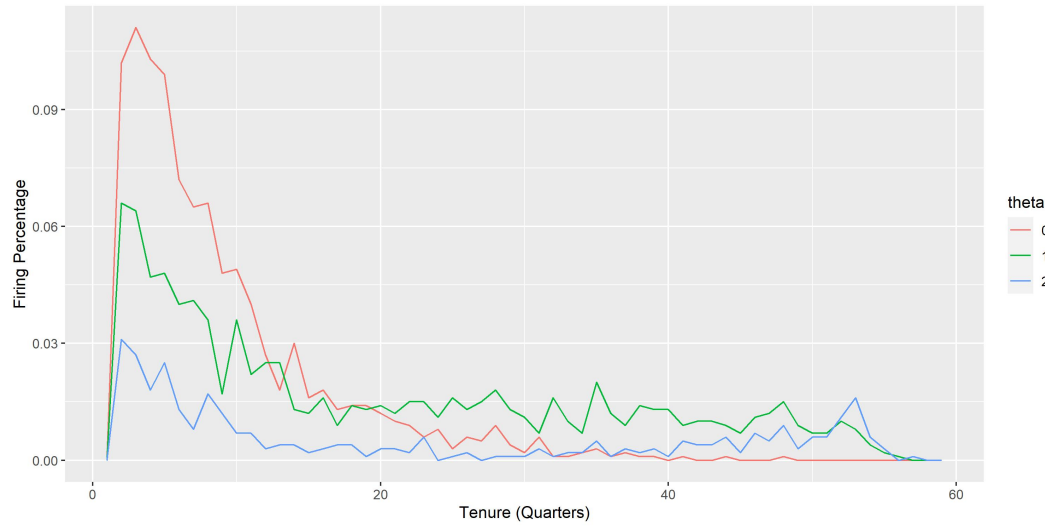


Figure 1.9: Unconditional firing percentage of CEO with skill $\theta = 0, 1, 2$ when the CEO is allowed to engage in short-term action.

notice that not only the lower skilled CEO but also the higher skilled CEO gets fired less at the early stages of their tenure. This implies that allowing the CEO to engage in short-term action can prevent the board from accidentally firing high skilled CEOs due to some bad luck in the early stages in their tenure. This shows that CEO's playing some short-term action can sometimes be beneficial for the firm.

There are two main points to take away from this section. One is that the lower skilled CEO is more likely to engage in short-term action. This could lead to low skilled CEO to survive longer than he should which is harmful to firm value. In addition the CEO tends to focus more on the short-term when they are close to getting fired. This relates to the first point since the lower skilled CEO is closer to getting fired than the higher skilled CEO. This shows that the threat of getting replaced induces the CEO to focus more on the short-term performance. Even though it is essential to replace the low skilled CEO to maximize firm value, the firm could suffer from CEO focusing more on short-term because of the threat of getting replaced.

1.3.4 Comparative Statics

In this section I illustrate how the model prediction changes with change in parameters. I illustrate how the results change when there is a change in persistence parameter ϕ , the standard deviation of the profits σ_η , the cost of firing C_f and the effectiveness of the short-term action C_a . Parameters take the value $\beta = 0.99$, $\mu_0 = 1$, $\sigma_0 = 1$, $\phi = 0.8$, $C_f = 8$, $\sigma_\eta = 3$, $C_a = 4$ if not mentioned.

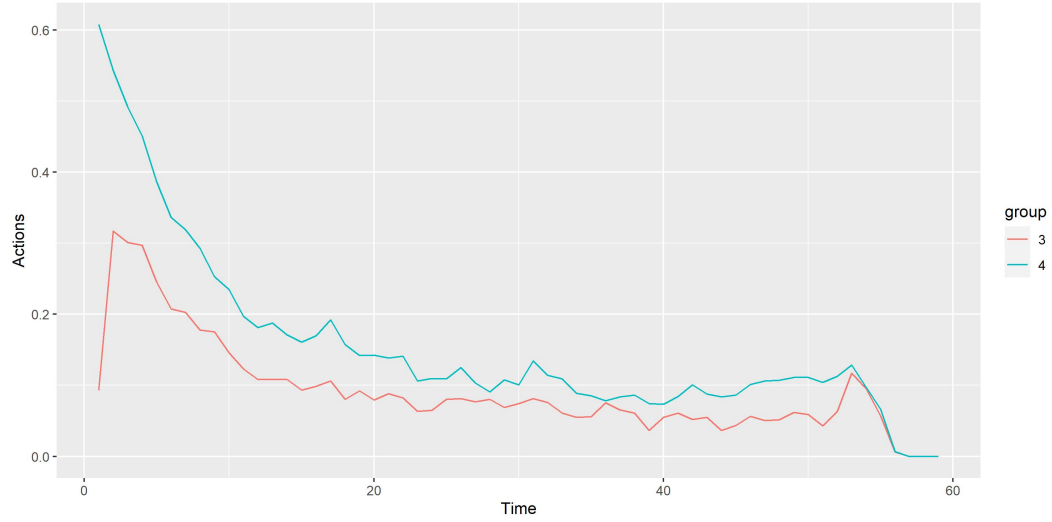


Figure 1.10: Average actions taken by the CEO for each $C_a = 3, 4$. Average is taken among θ and $\{\eta_t\}_{t=1}^{60}$.

Figure 1.10 shows the average actions taken by the CEO versus CEO tenure for two $C_a = 3, 4$. The average is taken among θ and $\{\eta_t\}_{t=1}^{60}$. Naturally increasing the effectiveness of the short-term action will increase the short-term action taken by the CEO.

Figure 1.11 shows the average actions taken by the CEO for two for two $C_f = 8, 10$. Increasing the cost of firing will decrease the firing threshold since it becomes less worth to fire a CEO. Therefore the CEO's job becomes more secure which allows them focus more on long-term.

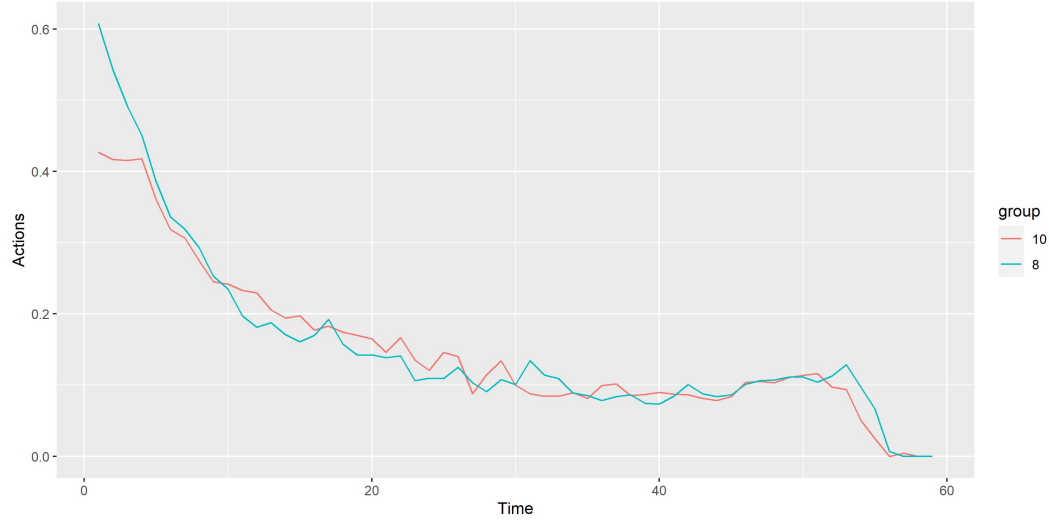


Figure 1.11: Average actions taken by the CEO for each $C_f = 8, 10$. Average is taken among θ and $\{\eta_t\}_{t=1}^{60}$.

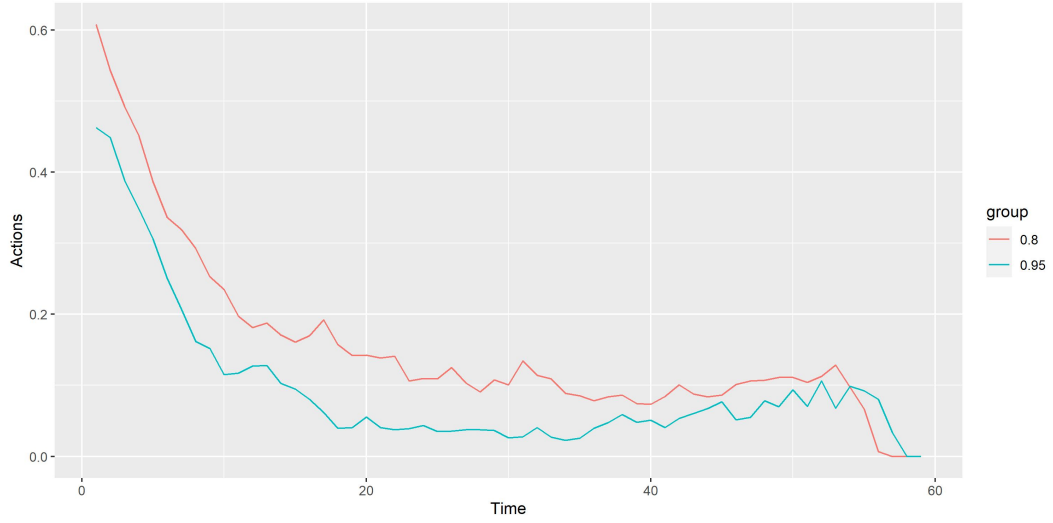


Figure 1.12: Average actions taken by the CEO for each $\phi = 0.8, 0.95$. Average is taken among θ and $\{\eta_t\}_{t=1}^{60}$.

Figure 1.12 shows average actions taken by the CEO for two $\phi = 0.8, 0.95$. Increasing the persistence parameter ϕ will have two effects. First as shown in the previous section it increases the firing threshold since having a low skilled CEO is more punishing to the firm. This should increase the incentive to play more short-term action. Second, it makes the

short-term action more harmful for firm value. This makes the disutility that the CEO incur from decline in share value caused short-term action increases. In this example, the second effect mostly outweighs the first effect.

1.3.5 The Cost of Short-term Action

In this section I analyze how CEO's action affects firm value.

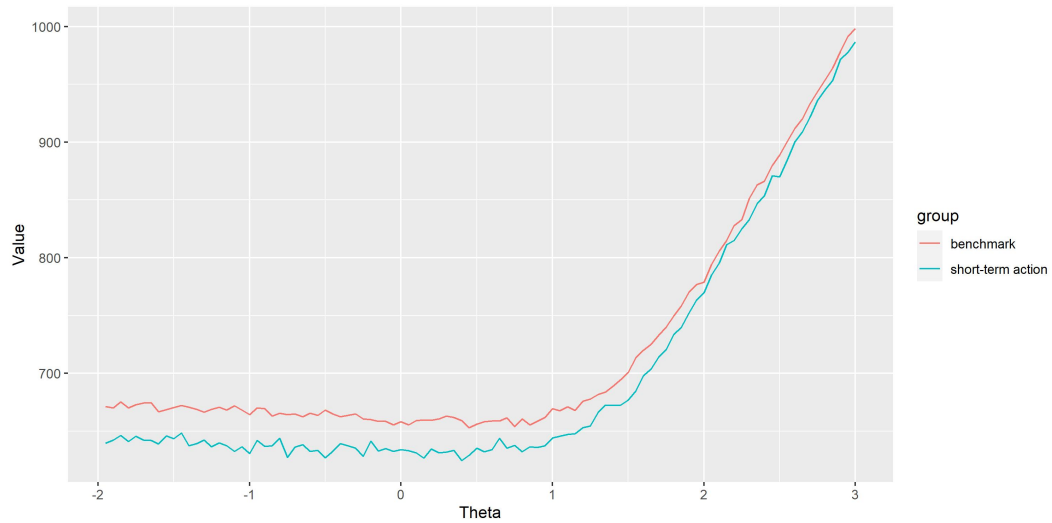


Figure 1.13: Value of the firm

Figure 1.13 plots the value of the firm versus the CEO skill θ . Recall that the value of the firm is given by equation (1.2.1) which is the discounted sum of profits.

Notice that the value of the firm is more or less flat below $\theta = 1$. This is a little unintuitive since it is natural to think that hiring a higher skilled CEO is always better for the firm, hence the value should be monotonically increasing in CEO skill. However, the lower skilled CEOs are more likely to get fired at the early stage of their tenure which allows the firm to hire a new CEO. For example although the CEO with skill $\theta = 0.5$ will generate more discounted profit over any fixed amount of period than the CEO with $\theta = 0$, since the CEO with skill $\theta = 0$ is more likely to get fired earlier than the CEO with skill $\theta = 0.5$, hiring the CEO with

skill $\theta = 0$ could be better for the firm. Having more or less a flat value function below $\theta = 1$ means that benefit from having higher skill is cancelled out from the effect mentioned above for skill $\theta < 1$. This shows that risk of hiring a low skilled CEO is fairly low since the board has the option to fire the CEO. This also relates to the previous result that the board prefers CEOs with higher variances since the downward risk is capped.

Obviously value of the firm is higher in the benchmark case since then short-term action is not allowed. However, notice that the difference between the benchmark case and the model case is small when the high skilled CEO is hired. There are at least 2 reasons for this. First, as shown in the previous section, a high skilled CEO less likely to engage in short-term action. Thus the difference between the benchmark case and the model case becomes small. Another reason is that some level of short-term action could be beneficial for the firm. This is because CEO playing some level of short-term action could prevent the board from accidentally firing the high skilled CEO. Especially in the early stages of CEO tenure, even a high skilled CEO are at risk of losing his job due to bad luck. So in the early stages of the CEO's tenure, even the highly skilled CEO may engage in some level of short-term action which will protect them from getting fired. Even though this action will harm the health of the firm, it could be beneficial for the firm as well since the firm is more likely to keep the highly skilled CEO.

The CEO engaging in short-term action has at least two effects on firm value. First, the CEO engaging in short-term action deters firms future profitability. As mentioned before, engaging in short-term action will change the discounted sum of profits by $C_a \cdot a_t - \frac{a_t}{1 - \beta \cdot \phi}$. Assumption 1 ensures that this value is negative. Hence, engaging in short-term action has a negative effect on firm value. Second, the CEO engaging in short-term action prevents the board from updating the CEO's skill properly. Since the board uses realized profits to update about the CEO's skill based on the assumption that $a_t = 0$ for t , the CEO engaging in

short-term action can disrupt the boards beliefs about the CEO's skill. In order to characterize these 2 effects I compare the value of the firm between the benchmark case ($a_t = 0$ for all t) and the model case which is the case where the CEO is allowed to engage in short-term action. The table below shows the value of the firm at time 0 when $h_0 = 0$ for 3 cases for different parameters.

Table 1.1: Effects of Short-term Actions on Firm Value

	V1(model)	V2(benchmark)	V3(no firing)	Loss1	Loss2	V2-V1
$C_a = 3, C_f = 8$	692	714	480	22.3	0.074	22.4
$C_a = 4, C_f = 8$	688	714	480	19.3	6.98	26.3
$C_a = 3, C_f = 10$	671	691	480	19.6	0.38	20.0
$C_a = 4, C_f = 10$	668	691	480	16.8	6.73	23.5

- V1 denotes the value of the firm in the model case where the CEO can engage in positive short-term action.
- V2 denotes the value of the firm in the benchmark case where the CEO always plays $a_t = 0$.
- V3 denotes the value of the firm in the no firing case where the board never fires the CEO and therefore CEO always survives for 60 periods.
- Loss 1 is the first effect mentioned above which is the direct loss from the short-term profit.
- Loss 2 is the calculated as $V2 - V1 - \text{Loss1}$. This describes second effect mentioned above which is the loss caused from the board not able to update the CEO skill properly.

Table shows that Loss 1 is more significant compared to Loss 2. This relates to the previous point where manipulating the firm's belief is sometime beneficial to the firm since it can prevent the firm from firing a high skilled CEO. As we can see from figure 13, the value of the

firm is almost flat at lower levels of θ but starts increasing once θ reach a certain level. This implies that it is extremely important to not accidentally fire a high skilled CEO. So when the firm is hiring a high skilled CEO, the CEO boosting the board's beliefs about his skill is beneficial for the firm. When the firm is hiring a low skilled CEO, the CEO boosting the board's beliefs about his skill may prevent the firm from firing the CEO, which leads to a loss in firm value. However, even if the low skilled CEO try to manipulate the board's belief, the board will eventually find out that the CEO is low skilled and he will get fired in a few periods. On the other hand, accidentally firing a high skilled CEO means that the firm have lost out on the value that the high skilled CEO could have created for 60 periods. Hence, the loss from keeping the low skilled CEO for an extra few periods is marginal compared to the potential loss from accidentally firing a high skilled CEO. Therefore, even though the belief manipulation from the low skilled CEO happens more often than the belief manipulation from the high skilled CEO, the loss from the belief manipulation becomes marginal. However, notice that once the effectiveness of the short-term action C_a increases from 3 to 4, Loss 1 declines and Loss 2 becomes more significant. Loss 1 decreasing is intuitive since the short-term action is more effective the direct loss from engaging in short-term action which is given by $C_a \cdot a_t - \frac{a_t}{1 - \beta \cdot \phi}$ will decrease. However, as shown in Figure 10, increasing C_a will result in CEO focusing more on the short-term which increases a_t . Also increase in C_a means that board's belief about the CEO skill is disrupted more by the short-term action a_t . This leads to low skilled CEO surviving much longer than they should, which increases Loss 2 as a result.

1.4 Conclusion

This paper develops dynamic model in which the CEOs can manipulate their performance measure at the expense of firm value. Instead of focusing on the conflict of interest between

principal and the agent regarding the agent's effort level, I focus on the conflict of interest regarding the allocation of the agent's effort between the short-term and the long-term. The model features learning where the board learns about the CEO's skill based on performance and makes a decision whether to keep or fire the CEO. The CEO can perform an action which boosts the short-term profit but deters long-term profitability of the firm. While the board only cares about the value of the firm, the CEO cares both about his job security and the value of the firm. By solving the model numerically I analyze the CEO's behavior and how the conflict of interest between the board and the CEO affect the value of the firm. I find that CEO with lower skill are more likely to focus on the short-term performance. Also the CEO who are close to getting fired tend to focus more on short-term performance. The CEO focusing on short-term performance not only deteriorates the firms future profitability but also prevents firms from making optimal firing decisions. Furthermore, I find that inefficiency caused by firms not able to update the skill of the CEO is smaller compared to the inefficiency caused by direct effect from CEO sacrificing long-term profitability for the sake of short-term performance.

Bibliography

- [1] Bebchuk, Lucian Arye, and Lars A. Stole. "Do Short-Term Objectives Lead to Under or Over investment in Long-Term Projects?." *The Journal of Finance* 48.2 (1993): 719-730.
- [2] Brickley, James A., James S. Linck, and Jeffrey L. Coles. "What happens to CEOs after they retire? New evidence on career concerns, horizon problems, and CEO incentives." *Journal of Financial Economics* 52.3 (1999): 341-377.
- [3] Campbell, Tim S., and Anthony M. Marino. "Myopic investment decisions and competitive labor markets." *International Economic Review* (1994): 855-875.
- [4] Dahiya, Sandeep, and David Yermack. "You can't take it with you: Sunset provisions for equity compensation when managers retire, resign, or die." *Journal of Corporate Finance* 14.5 (2008): 499-511.
- [5] DeMarzo, Peter M., and Yuliy Sannikov. "Learning, termination, and payout policy in dynamic incentive contracts." *The Review of Economic Studies* 84.1 (2016): 182-236.
- [6] Edmans, Alex, Vivian W. Fang, and Katharina A. Lewellen. Equity vesting and managerial myopia. No. w19407. National Bureau of Economic Research, 2013.
- [7] Eisfeldt, Andrea L., and Camelia M. Kuhnen. "CEO turnover in a competitive assignment framework." *Journal of Financial Economics* 109.2 (2013): 351-372.

- [8] Gopalan, R., Milbourn, T., Song, F., and Thakor, A. V. (2014). Duration of executive compensation. *The Journal of Finance*, 69(6), 2777-2817.
- [9] Graham, John R., Campbell R. Harvey, and Shiva Rajgopal. "The economic implications of corporate financial reporting." *Journal of accounting and economics* 40.1-3 (2005): 3-73.
- [10] Laux, Volker. "Stock option vesting conditions, CEO turnover, and myopic investment." *Journal of Financial Economics* 106.3 (2012): 513-526.
- [11] Marinovic, Iván, and Felipe Varas. CEO Horizon, Optimal Pay Duration and the Escalation of Short-termism. Working Paper, 2017.
- [12] Stein, Jeremy C. "Efficient capital markets, inefficient firms: A model of myopic corporate behavior." *The Quarterly Journal of Economics* (1989): 655-669.
- [13] Taylor, Lucian A. "Why are CEOs rarely fired? Evidence from structural estimation." *The Journal of Finance* 65.6 (2010): 2051-2087

Chapter 2

On Using Linear Quantile Regressions for Causal Inference

2.1 Introduction

Is it appropriate to use the linear quantile regression (QR) for causal inference? To answer this question, we study relationships among the linear QR, the conditional quantile function (CQF), and general structural functions. First, we show that the slope parameter of the linear QR represents a weighted average of the local slopes of the CQF. Second, we cite an existing result demonstrating that the slope of the CQF represents a weighted average of the structural partial effects. Third, by chaining these two arguments, we show that the slope parameter of the linear QR represents a weighted average of the structural partial effects. Our results imply that the linear QR can be used for structural causal inference in the presence of nonlinearity and multiple unobserved heterogeneity.

Endogeneity and misspecification could be two major obstacles in using the linear QR for

causal inference. Since an extensive body of the econometrics literature has solved the endogeneity issue in various contexts,¹ we assume exogeneity and focus on the misspecification problem throughout this paper. The linear QR generally misspecifies the true causal structure in two ways. First, the linearity of the linear QR fails to capture the nonlinearity of general structural functions.² Second, the monotonicity of the linear QR with respect to a scalar latent variable fails to capture the non-monotonicity of general structural functions with respect to generally multivariate unobservables. Despite these two sources of potential misspecification, what we develop in this paper reconciles the linear QR with the true causal structure.

It is a well-known feature of the ordinary least squares (OLS) method that it minimizes the average squared distance between the linear regression function and the conditional expectation function (CEF). In fact, the OLS does not only minimize the fit. Yitzhaki (1996) and Angrist and Krueger (1999) show that the OLS slope coefficient under misspecification can be explicitly written as weighted average derivatives of the CEF. This result is viewed by empirical researchers to be useful for an interpretation of regression coefficients.³ One contribution of this paper is to provide a generalization of this weighted-average-derivatives interpretation for arbitrary linear functions that minimize weighted mean squared distances. Our result applies to the linear QR in particular, since Angrist, Chernozhukov, and Fernández-Val (2006) show that the linear QR parameters minimize a weighted average of squared distances between the linear function and the CQF.⁴

In Section 2.2, we show that the slope parameter of the linear QR equals a weighted average of the local slopes of the CQF. In Section 2.4, we use the existing result that the slope of the CQF equals a weighted average of the structural partial effects. We provide a framework

¹See Chernozhukov and Hansen (2005) and many subsequent papers.

²In addition to causal interpretation, the linear misspecification of the linear QR can cause the problem of quantile crossing. Chernozhukov, Fernández-Val, and Galichon (2010) propose rearrangement as a solution.

³Furthermore, even under endogeneity, the two-stage least squares have the weighted-average interpretation for discrete treatment effects (Angrist and Imbens, 1995).

⁴Lee (2014) provides an alternative measure of fit for the linear QR parameters.

for claiming the two auxiliary results together to show that the slope parameter of the linear QR equals a weighted average of the structural partial effects, even if the underlying structural function is nonlinear in the explanatory variable and/or is non-monotone with respect to possibly multivariate latent variable(s). We therefore conclude in Section 2.6 that linear QR (Koenker and Bassett, 1978)⁵ is a useful tool for causal inference in the presence of unobserved heterogeneity, even if it misspecifies the true structure both in terms of nonlinearity and non-monotonicity. A couple of extensions to the baseline result are presented in Section 2.3, and we also argue that the same conclusion applies to the linear regression in Section 2.5 as well as the linear QR.

2.2 Relationship between Linear and Nonlinear Functions

2.2.1 General Setting

We first introduce basic definitions and fix relevant notations. Let Y and X denote dependent and independent variables supported on $\mathcal{Y} \subset \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}$, respectively. Let $\mathcal{B}(\mathcal{X})$ denote the Borel sigma algebra on \mathcal{X} . First, we present an auxiliary lemma that a nonlinear function and a linear function related by a minimum weighted mean squared distance have at least two intersection points. Let $M : \mathcal{X} \rightarrow \mathbb{R}$ be a general function and $L : \mathcal{X} \rightarrow \mathbb{R}$ be an affine function specified by the parametric form $L(x) = \alpha + \beta x$. We state the following assumption for these two functions and the distribution of X .

Assumption 3. (i) $E[M(X)^2]$ and $E[X^2]$ are finite. (ii) (α, β) uniquely solves $\min_{(a,b)} E[\omega(X) \cdot (M(X) - a - bX)^2]$ for some weight function $\omega : \mathcal{X} \rightarrow \mathbb{R}$, where $\omega(x) \geq 0$ a.e. (iii) $\omega(X)$ is bounded a.s. (iv) M is continuous. (v) \mathcal{X} is convex. (vi) X is continuously distributed.

In this assumption, the linear function L is characterized as the unique weighted least

⁵See also Koenker and Hallock (2001), Koenker (2005), and Chernozhukov and Hansen (2013) for surveys.

squares approximation to the nonlinear function M . This applies to the relation between the linear regression function L and the CEF M , for example. In addition, this also applies to the relation between the linear QR function L and the CQF M by Angrist, Chernozhukov, and Fernández-Val (2006). The current section proceeds with a general setting, and we will branch into these two concrete examples in Sections 2.2.3 and 2.2.4.

We first obtain the following auxiliary lemma with Assumption 3.

Lemma 1 (Existence of Two Intersection Points). *If Assumption (3) is true, then M and L intersect at least at two points in \mathcal{X} .*

A proof of this lemma is provided in Section B.1 in the appendix. We also present an intuitive illustration of this lemma in Section 2.2.2, describing how the minimum weighted mean-squares characterization leads to the existence of two intersection points. Applying Lemma 1, we next claim that the slope β of L is a weighted average of the local slopes M' of M . To this end, we invoke an additional assumption that ensures that this derivative M' of interest exists.

Assumption 4. *M is continuously differentiable.*

An application of Lemma 1 and the Fundamental Theorem of Calculus yields the following result.

Theorem 1. *If Assumptions 3 and 4 are true, then there exist $x_1, x_2 \in \mathcal{X}$ with $x_1 < x_2$ such that*

$$\beta = E \left[w_{x_1, x_2}(X) \cdot M'(X) \mid x_1 \leq X \leq x_2 \right]$$

holds, where the weight function w_{x_1, x_2} is given by

$$w_{x_1, x_2}(x) := \frac{\frac{1}{f_X(x)}}{\int_{x_1}^{x_2} \frac{1}{f_X(\xi)} dF_X(\xi)} \quad \text{for all } x \in \mathcal{X}.$$

A proof of this theorem is provided in Section B.2 in the appendix, but it is straightforward from Lemma 1. This result characterizes β as a weighted average of M' . Furthermore, note that the weight function w_{x_1, x_2} is strictly positive on $[x_1, x_2]$. In other words, the slope β of L represents a strict convex combination of the local slopes M' of M within an interval $[x_1, x_2] \subset \mathcal{X}$, which is guaranteed to be non-empty.

To mathematically formalize the weighted-average interpretation, we define the function $\lambda_{x_1, x_2} : \mathcal{B}(\mathcal{X}) \rightarrow \mathbb{R}$ by

$$\lambda_{x_1, x_2}(T) = \int_{T \cap [x_1, x_2]} w_{x_1, x_2}(x) dF_X(x) \quad \text{for each } T \in \mathcal{B}(\mathcal{X}). \quad (2.2.1)$$

Note that this λ_{x_1, x_2} can be shown to be a probability measure. With this notation, the conclusion of Theorem 1 can be restated as

$$\beta = \mathbb{E}_{\lambda_{x_1, x_2}} [M'(X)]. \quad (2.2.2)$$

In other words, the slope β of L is the mean of the local slopes M' of M with respect to the probability measure λ_{x_1, x_2} defined in (2.2.1).

A remaining question is whether there can exist more than two intersection points. In general, there may exist more than two. For example, in the extreme case where M is also an affine function, the minimum weighted mean squared distance restriction necessarily leads to $L = M$, and there exists a continuum of intersection points. Note that, in this affine case, one can take any pair of points, x_1 and x_2 , from \mathcal{X} to form the weight function w_{x_1, x_2} defined in Theorem 1, since $\beta = L' = M'$ holds globally. In general, when there exist more than two intersection points between L and M , we can choose any pair of those intersection points, x_1 and x_2 , to form the weight function w_{x_1, x_2} . In this sense, the weight function need not be unique.

2.2.2 Intuition

We provide an intuitive illustration for Lemma 1, where we claim that there exist at least two intersection points of M and L in \mathcal{X} . Since L minimizes the weighted mean squared distance from M , no other affine function can have a strictly smaller w -weighted mean squared distance from the function M than the affine function L . This restriction plays the major role in our theoretical and intuitive arguments.

By way of contradiction, assume that L and M have less than two intersection points in \mathcal{X} . We can then branch into three representative cases, illustrated by (a)–(c) in Figure 2.1. Panel (a) shows a case where there is no intersection point. In this case, either $L - M > 0$ or $L - M < 0$ is true by the continuity of M . If $L - M < 0$ like the graph, then shifting L upward to produce \tilde{L} would achieve a smaller weighted mean squared distance between M and \tilde{L} than between M and L , contradicting the premise that L minimizes the weighted mean squared distance from M . Thus, we rule out case (a) of zero intersection point.

Panels (b) and (c) show cases where there is only one intersection point. In these cases, rotating and/or shifting L to produce \tilde{L} would achieve a smaller weighted mean squared distance between M and \tilde{L} than between M and L , contradicting the characterization of L as the minimizer of the weighted mean squared distance from M . Thus, we rule out cases (b) and (c) of one intersection point.

While this illustration is intuitive, establishing the existence of such a rotation and/or shift to achieve a strictly smaller weighted mean squared distance is not necessarily a trivial problem, except in case (a). We refer readers to Section B.1 in the appendix for mathematical details of this illustration.

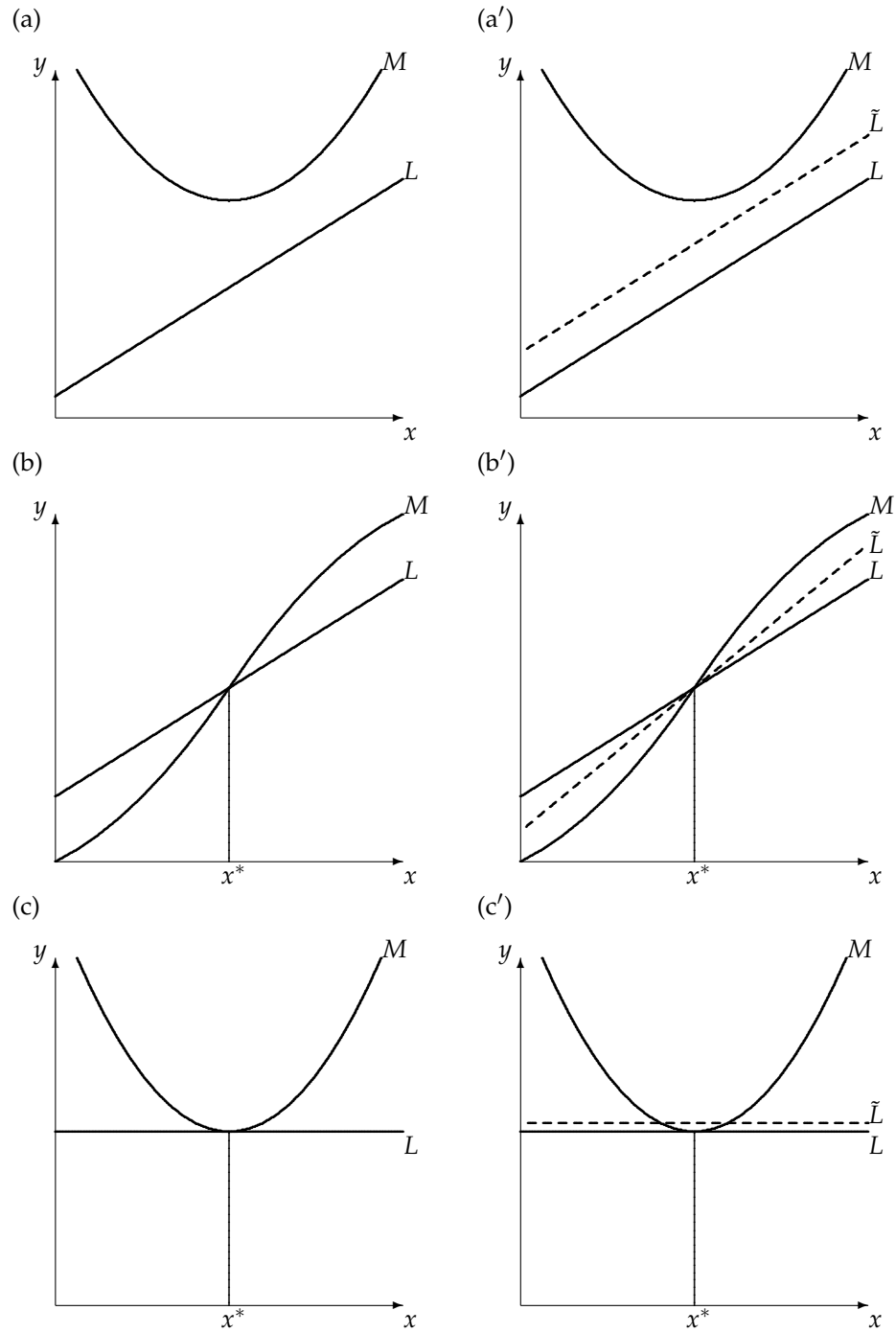


Figure 2.1: Illustrations of why M and L must have at least two intersection points. Panel (a) shows a case where there is no intersection point. Panels (b) and (c) show cases where there is only one intersection point. In each of these cases, moving L to \tilde{L} may reduce the weighted mean squared distance from M , and hence L cannot minimize the weighted mean squared distance.

2.2.3 Linear QR and CQF

We demonstrate an application of Theorem 1 to the case of quantile regressions. For a given $\tau \in (0, 1)$, the τ -th CQF $Q_\tau : \mathcal{X} \rightarrow \mathcal{Y}$ is defined by

$$Q_\tau(x) = \inf\{y \mid F_{Y|X}(y \mid x) \geq \tau\},$$

where $F_{Y|X}(y \mid x)$ denotes the conditional cumulative distribution function of Y given X , which is assumed to be regularly defined throughout the paper. The τ -th linear QR L_τ provides the best linear prediction of Y given X under the loss function $\rho_\tau(u) : \mathbb{R} \rightarrow \mathbb{R}$ defined by $\rho_\tau(u) = (\tau - \mathbb{1}(u \leq 0)) \cdot u$, called the check function. In other words, $L_\tau(x) = \alpha(\tau) + \beta(\tau)x$ where $(\alpha(\tau), \beta(\tau))$ satisfies

$$(\alpha(\tau), \beta(\tau)) \in \arg \min_{(\alpha, \beta) \in \Theta} E[\rho_\tau(Y - \alpha - \beta X)]$$

for some parameter set $\Theta \subset \mathbb{R}^2$. Note that we have $Q_\tau \neq L_\tau$ in general.

Angrist, Chernozhukov, and Fernández-Val (ACF, 2006; Theorem 2) show that, under the following assumption, the linear QR minimizes a weighted mean squared distance from the CQF. We state their results below.

Assumption 5 (ACF, 2006). (i) The conditional density $f_{Y|X}(\cdot \mid X)$ exists and is bounded a.s. (ii) $E[Y]$, $E[Q_\tau(X)^2]$, and EX^2 are finite. (iii) $(\alpha(\tau), \beta(\tau))$ uniquely solves $\min_{(\alpha, \beta)} E[\rho_\tau(Y - \alpha - \beta X)]$.

Theorem 2 (ACF, 2006; Theorem 2). If Assumption 5 is true, then $(\alpha(\tau), \beta(\tau))$ uniquely solves the weighted least squares problem

$$\min_{\alpha, \beta} E[\bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)) \cdot (L_\tau - Q_\tau)^2(X, \alpha, \beta)], \quad (2.2.3)$$

where $\bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau))$ is defined by

$$\bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) := \frac{1}{2} \int_0^1 f_{Y|X}(uL_\tau(x) + (1-u)Q_\tau(x) \mid x) du.$$

To claim that the slope $\beta(\tau)$ of the τ -th linear QR L_τ is a weighted average of the slopes Q'_τ of the CQF, we now invoke an additional assumption that the support of X is convex and that the derivative Q'_τ of interest exists.

Assumption 6. (i) \mathcal{X} is convex. (ii) Q_τ is continuously differentiable. (iii) \mathcal{X} is continuously distributed.

Note that Assumptions 5 and 6 imply Assumption 3 by Theorem 2. Note also that Assumption 6 (ii) implies Assumption 4. Combining Theorems 1 and 2 together, therefore, we get the following corollary that the slope $\beta(\tau)$ of the τ -th linear QR L_τ is a weighted average of the slopes Q'_τ of the CQF.

Corollary 1 (Weighted Average: CQF). *If Assumptions 5 and 6 are true, then there exist $x_1, x_2 \in \mathcal{X}$ with $x_1 < x_2$ such that*

$$\beta(\tau) = E [w_{x_1, x_2}(X) \cdot Q'_\tau(X) \mid x_1 \leq X \leq x_2]$$

holds, where the weight function w_{x_1, x_2} is defined in Theorem 1.

As in the general case, this result can be formally interpreted as a weighted average based on the probability measure defined in (2.2.1). Specifically, the conclusion of Corollary 1 can be restated as

$$\beta(\tau) = E_{\lambda_{x_1, x_2}} [Q'_\tau(X)] . \quad (2.2.4)$$

2.2.4 Linear Regressions and CEF

While the main focus of this paper is on quantile regressions, we demonstrate in the current subsection an application of Theorem 1 to the case of mean regressions. The linear regression L_m is a linear function which gives the minimum mean-squared-errors prediction of Y given

X , i.e., $L_m(x) = \alpha + \beta x$ where

$$(\alpha, \beta) \in \arg \min_{(a,b)} E[(Y - a - bX)^2].$$

In addition, the linear regression also provides the minimum mean-squared-distance approximation to the CEF, i.e., L_m also satisfies

$$(\alpha, \beta) \in \arg \min_{(a,b)} E[(m(X) - a - bX)^2],$$

where $m : \mathcal{X} \rightarrow \mathbb{R}$ denotes the CEF. The approximation properties of the linear regression have been emphasized by White (1980), Chamberlain (1984), Goldberger (1991), and Angrist and Krueger (1999), among others. To apply Theorem 1 to the mean regressions, we state the following assumptions.

Assumption 7. (i) $E[Y^2]$ and $E[X^2]$ are finite. (ii) (α, β) uniquely solves $\min_{(a,b)} E[(Y - a - bX)^2]$. (iii) \mathcal{X} is convex. (iv) X is continuously distributed. (v) m is continuously differentiable.

Note that, if (α, β) uniquely solves $\min_{(a,b)} E[(Y - a - bX)^2]$, then (α, β) also uniquely solves $\min_{(a,b)} E[(m(X) - a - bX)^2]$. Therefore, Assumption 7 implies Assumptions 3 and 4, and we consequently obtain from Theorem 1 the following corollary that the slope β of the linear regression L_m is a weighted average of the slopes m' of the CEF.

Corollary 2 (Weighted Average: CEF). *If Assumption 7 is true, then there exist $x_1, x_2 \in \mathcal{X}$ with $x_1 < x_2$ such that*

$$\beta = E [w_{x_1, x_2}(X) \cdot m'(X) | x_1 \leq X \leq x_2]$$

holds, where the weight function w_{x_1, x_2} is defined in Theorem 1.

As in the general case, this result can be formally interpreted as a weighted average based

on the probability measure defined in (2.2.1). The conclusion of Corollary 2 can be restated as

$$\beta = E_{\lambda_{x_1, x_2}} [m'(X)] .$$

We conclude this section with a remark about existing results that derive weighted average interpretations for the OLS slope parameter. Yitzhaki (1996; Proposition 2) shows that the OLS slope parameter can be written as a weighted average of the local slopes of the CEF with the weight function w_Y defined by

$$w_Y(x) = \frac{1}{\text{Var}(X)} [E[X] \cdot F_X(x) - \Theta_X(x)] = \frac{F_X(x)}{\text{Var}(X)} (E[X] - E[X|X \leq x]),$$

where $\Theta_X(x) = \int_{-\infty}^x tf(t)dt = F_X(x)E[X|X \leq x]$. Likewise, Angrist and Krueger (1999; page 1311) show that the OLS slope parameter can be written as

$$\beta = E \left[\frac{X^2}{E[X^2]} m'(\zeta(X)) \right] ,$$

where $\zeta(x)$ derives from the mean value expansion around 0, and is thus between 0 and x . If ζ is invertible and continuously differentiable, then the resultant weight function w_{AK} can be defined by

$$w_{AK}(x) = \frac{\zeta^{-1}(x)^2 \cdot f_X(\zeta^{-1}(x))}{\zeta'(\zeta^{-1}(x)) \cdot f_X(x) \cdot E[X^2]} .$$

The weight function w_Y of Yitzhaki, the weight function w_{AK} of Angrist and Krueger, and our weight function w_{x_1, x_2} defined in Theorem 1 are different from each other. The linear regression slope parameter can thus admit multiple weights for the weighted-average interpretation. Each of the different weights can provide a unique way of interpretation. For example, w_Y has an advantage in that it is strictly positive on the support \mathcal{X} , while w_{x_1, x_2} has a truncated support $[x_1, x_2] \subset \mathcal{X}$. Thus, Yitzhaki's result is useful for interpreting the slope parameter as a weighted average over the entire population. A similar remark applies to the

weight of Angrist and Krueger, except for the zero weight at the point $x = 0$. On the other hand, w_{x_1, x_2} allows for a more intuitive interpretation as the inverse probability weighting. Thus, our result is useful for interpreting the slope parameter as a standardized average over a subpopulation. Because all the weights are true anyway, empirical practitioners can take advantage of all the convenient interpretations.

2.3 Extensions

This section presents a couple of extensions to the baseline result. First, we demonstrate that our basic idea also applies to the case of discrete regressors. Second, we demonstrate that our result also extends to the case of multivariate regressors.

2.3.1 The Result for the Case of Discrete Regressors

The main result (Theorem 1) was obtained assuming that the regressor X is continuously distributed. In the current section, we argue that a similar idea applies to the case of a discrete regressor, which is very common in economic applications.⁶ As we move from a continuous distribution to a discrete distribution, we also change the object of interest from derivatives to differences. The following short-hand notations are introduced.

$$\beta = \Delta L(x) \quad := \quad L(x') - L(x)$$

$$\Delta M(x) \quad := \quad M(x') - M(x)$$

where $x' = \min\{\tilde{x} \in \mathcal{X} | \tilde{x} > x\}$. As in the continuous case, we first prove an auxiliary lemma under the following assumption.

Assumption 8. (i) $E[M(X)^2]$ and $E[X^2]$ are finite. (ii) (α, β) uniquely solves $\min_{(a,b)} E[\omega(X) \cdot$

⁶We thank K. Kato for suggesting this extension.

$(M(X) - a - bX)^2]$. (iii) $\omega(X)$ is bounded for all $x \in \mathcal{X}$. (iv) $\mathcal{X} \subset \mathbb{Z}$ and $|\mathcal{X}| \geq 2$.

This assumption is analogous to Assumption 3, which we made for the case of continuous X . Compared to that assumption, the current assumption drops the continuity of M and the continuous distribution of X . Instead, it adds the restriction that \mathcal{X} is integer-valued, although it can be relaxed to arbitrary discrete supports with isolated points. Under this assumption, we obtain the following auxiliary lemma.

Lemma 2. *If Assumption 8 is true, then there exist two distinct points $x_1, x_2 \in \mathcal{X}$ such that $(\Delta L(x_1) - \Delta M(x_1)) \cdot (\Delta L(x_2) - \Delta M(x_2)) \leq 0$ holds.*

A proof is found in Section B.3 in the appendix. It is proved analogously to the logic used in the proof of Lemma 1, which we developed for the case of continuous X . This lemma claims that there is some point $x_1 \in \mathcal{X}$ at which $\beta = \Delta L(x_1) \geq \Delta M(x_1)$ is true and another point $x_2 \in \mathcal{X}$ at which $\beta = \Delta L(x_2) \leq \Delta M(x_2)$ is true. In other words, the slope parameter β of the linear function is bounded from both above and below by the differences ΔM . This property immediately implies that β can be written as a convex combination of $\{\Delta M(x) \mid x \in \mathcal{X}\}$, as formally stated in the following theorem.

Theorem 3. *If Assumption 8 is true, then*

$$\beta(\tau) = E[\tilde{w}(X) \cdot \Delta M(X)]$$

holds for a non-negative weight function \tilde{w} such that $E[\tilde{w}(X)] = 1$.

A proof is provided in Section B.4 in the appendix. While our proof constructs a particular weight function w to establish the equality, such a weight function need not be unique.

Like the case of continuous regressors, we can apply Theorem 3 to quantile regressions in particular. Let $\Delta Q_\tau(x) := Q_\tau(x') - Q_\tau(x)$. The following corollary follows from Theorem 3.

Corollary 3. *If Assumption 5 and Assumption 8(iv) are true, then*

$$\beta(\tau) = E[\tilde{w}(X) \cdot \Delta Q_\tau(X)]$$

holds for a non-negative weight function \tilde{w} such that $E[\tilde{w}(X)] = 1$.

Likewise, we can apply Theorem 3 to mean regressions as well. Let $\Delta m(x) := m(x') - m(x)$. The following corollary follows from Theorem 3.

Corollary 4. *If Assumption 7(i)–(ii) and Assumption 8(iv) are true, then*

$$\beta(\tau) = E[\tilde{w}(X) \cdot \Delta m(X)]$$

holds for a non-negative weight function \tilde{w} such that $E[\tilde{w}(X)] = 1$.

2.3.2 The Result for the Case of Multivariate Regressors

The main result (Theorem 1) was obtained assuming that the regressor X is univariate. In the current section, we argue that a similar idea applies to the case of a multivariate regressor. To this goal, we make the following assumption.

Assumption 9. (i) $E[M(X)^2]$ and $E\|X\|^2$ are finite. (ii) $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^k$ uniquely solve $\min_{(a,b)} E[\omega(X) \cdot (M(X) - a - X'b)^2]$ for some weight function $\omega : \mathcal{X} \rightarrow \mathbb{R}$, where $\omega(x) \geq 0$ a.e. (iii) $\omega(X)$ is bounded a.s. (iv) M is twice continuously differentiable. (v) $\mathcal{X} = \mathbb{R}^k$. (vi) X is continuously distributed. (vii) There exists $C > 0$ such that $\frac{\partial^2 M(x)}{\partial x_i^2} > C$ for all $1 \leq i \leq k$ for all $x \in \mathcal{X}$.

Parts (i)–(vi) of this assumption are analogous to Assumption 3. In addition, part (vii) requires the nonlinear function M be convex.⁷ In econometrics, it is not unusual to impose

⁷We impose this additional assumption as we move from univariate X to multivariate X , because multivariate extension is not straightforward otherwise. Similar difficulties arose also for the weighted-average interpretations of the linear regressions by Yitzhaki (1996) and Angrist and Krueger (1999).

such shape restrictions, in light of the fact that many functions in economics must be convex or concave to satisfy important economic properties. We can substitute the assumption that M is concave, and arguments below then can be straightforwardly modified by reversing the inequalities. With x_{-1} denoting the $(k-1)$ -dimensional subvector $(x_2, x_3, \dots, x_k)'$ of x , we state the following two auxiliary lemmas that follow from the preceding assumption.

Lemma 3. *If Assumption 9 is true, then there exists $x \in \mathcal{X}$ such that $L(x) > M(x)$.*

Lemma 4. *Under Assumption 9, if there exists $x^* \in \mathcal{X}$ such that $L(x^*) > M(x^*)$, then $L(\cdot, x_{-1}^*)$ and $M(\cdot, x_{-1}^*)$ intersect at two distinct points.*

Proofs of Lemmas 3 and 4 are provided in Sections B.5 and B.6 in the appendix, respectively. Like the main result, an application of the Fundamental Theorem of Calculus together with these auxiliary lemmas yields the following weighted-average interpretation result.

Theorem 4. *If Assumption 9 is true, then*

$$\beta_1 = E \left[w_{\mathcal{D}}(X) \cdot \frac{\partial M(X)}{\partial X_1} \middle| X \in \mathcal{D} \right]$$

holds for the weight function $w_{\mathcal{D}}$ defined by

$$w_{\mathcal{D}}(x) := \frac{\frac{1}{x_1''(x_{-1}) - x_1'(x_{-1})} \cdot \frac{1}{f_X(x)}}{\int_{\xi \in \mathcal{D}_{-1}} \frac{1}{f_{X_{-1}}(\xi)} dF_{X_{-1}}(\xi)} \quad \text{for all } x \in \mathcal{X},$$

where

$$\mathcal{D} := \{x \in \mathcal{X} | L(x) > M(x)\},$$

$$\mathcal{D}_{-1} := \{x_{-1} \in \mathbb{R}^{k-1} | \exists x_1 \in \mathbb{R} \text{ s.t. } L(x) > M(x)\},$$

$$x_1''(x_{-1}) = \max\{x_1 \in \mathbb{R} | L(x_1, x_{-1}) = M(x_1, x_{-1})\}, \quad \text{and}$$

$$x_1'(x_{-1}) = \min\{x_1 \in \mathbb{R} | L(x_1, x_{-1}) = M(x_1, x_{-1})\}.$$

A proof is provided in Section B.7 in the appendix. We remark that this proof in particular shows that the weight function $w_{\mathcal{D}}$ is well-defined, as its denominator is strictly positive.

2.4 Linear QR and Structural Functions

A general class of structural functions can be expressed by a nonseparable function $g : \mathcal{X} \times \mathcal{U} \rightarrow \mathcal{Y}$ where $\mathcal{U} \subset \mathbb{R}^M$. The cumulative distribution function of U is denoted by F_U . Letting U denote an M -dimensional random vector of unobserved variables supported on \mathcal{U} , we can use g to write the relation among (Y, X, U) by

$$Y = g(X, U).$$

In the subsequent subsections, we explore relationships between the slopes of the linear QR and the structural partial effects $\partial g / \partial x$, where the latter object measures the *ceteris paribus* causal effects of X on Y .

2.4.1 Monotone Structural Functions

It is known that, if X is exogenous and g is monotone with respect to a scalar U , i.e., $M = 1$, then the CQF Q_{τ} can be used to represent the structural function g . Specifically, if $M = 1$ and $g(x, \cdot)$ is increasing for each $x \in \mathcal{X}$, then $Q_{\tau}(x) = g(x, u)$ holds for all $x \in \mathcal{X}$ where $\tau = F_U(u)$. In this case, Corollary 1 implies that the slope $\beta(\tau)$ of the τ -th linear QR L_{τ} identifies a weighted average of the structural partial effects $\partial g / \partial x$. We formally present this implication as the following corollary.

Corollary 5 (Weighted Average: Monotone Structural Function). *Suppose that $M = 1$, $g(x, \cdot)$ is monotone for each $x \in \mathcal{X}$, g is continuously differentiable, and X is exogenous. If Assumptions 5*

and 6 are true for $\tau := F_U(u)$ and $u \in \mathcal{U}$, then there exist $x_1, x_2 \in \mathcal{X}$ with $x_1 < x_2$ such that

$$\beta(\tau) = E \left[w_{x_1, x_2}(X) \cdot \frac{\partial g(X, u)}{\partial x} \middle| x_1 \leq X \leq x_2 \right]$$

holds, where the weight function w_{x_1, x_2} is defined in Theorem 1.

This result shows that the linear QR parameter $\beta(\tau)$ is useful for us to learn about the structural partial effects $\partial g / \partial x$ when the structural function g is monotone with respect to a scalar latent variable U . In the next subsection, we explore an interpretation of $\beta(\tau)$ while relaxing this monotonicity assumption.

2.4.2 General Structural Functions

We claim that the linear QR parameter can be still expressed as a weighted average of the structural partial effects even if the monotonicity assumption used in Corollary 5 is dropped. To this end, we combine our Corollary 1 with a result that exists in the literature, which connects the slope Q'_τ of the CQF to the structural partial effects $\partial g / \partial x$ under an arbitrary dimension M of the latent variables U .

We define the lower contour set $V(y, x) = \{u \in \mathbb{R}^M \mid g(x, u) \leq y\}$. Its boundary is denoted by $\partial V(y, x)$. Next, we define the algebra $\mathcal{B}(y, x) := \{S \cap \partial V(y, x) \mid S \in \mathcal{B}(\mathbb{R}^M)\}$ on $\partial V(y, x)$, where $\mathcal{B}(\mathbb{R}^M)$ is the Borel σ -algebra. Note that every element $S \in \mathcal{B}(y, x)$ is also a Borel set. Let m^M denote the Lebesgue measure on \mathbb{R}^M , and let H^{M-1} denote the $(M-1)$ -dimensional Hausdorff measure restricted to $(\partial V(y, x), \mathcal{B}(y, x))$. The velocities of the boundary $\partial V(y, x)$ at u with respect to a change in y and a change in x are denoted by $\partial v(y, x; u) / \partial y$ and $\partial v(y, x; u) / \partial x$, respectively. With Σ denoting an $(M-1)$ -dimensional rectangle, the boundary $\partial V(\cdot, \cdot)$ can be represented by a map $\Sigma \times \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}^M$, i.e., the $(M-1)$ -dimensional manifold $\partial V(y, x)$ can be parameterized through a map $\pi_{(y, x)} : \Sigma \rightarrow \partial V(y, x)$ for each $(y, x) \in \mathcal{Y} \times \mathcal{X}$. The velocity $\partial v(y, \cdot; \cdot) / \partial x$ of the boundary with respect

to a change in x can be represented by a map $\mathcal{X} \rightarrow L^1(\Sigma)$ for each $y \in \mathcal{Y}$, provided that the function $\partial v(y, x; \cdot)/\partial x : \partial V(y, x) \rightarrow \mathbb{R}^M$ is absolutely integrable with respect to the parameterization over Σ for each $(y, x) \in \mathcal{Y} \times \mathcal{X}$. Similarly, $\partial v(\cdot, x; \cdot)/\partial y$ can be represented by a map $\mathcal{Y} \rightarrow L^1(\Sigma)$ for each $x \in \mathcal{X}$, provided absolute integrability. We cite the following assumption and auxiliary lemma, which follows from an application of a conservation law for fluid mechanics – details and discussions can be found in Sasaki (2015).

Assumption 10. (i) g is continuously differentiable. (ii) $\|\nabla_u g(x, \cdot)\| \neq 0$ on $\partial V(y, x)$. (iii) The distribution of U is absolutely continuous with respect to m^M . (iv) $\int_{\partial V(y, x)} f_U(u) dH^{M-1}(u) > 0$. (v) $\partial V(y, \cdot) \in C^1(\Sigma \times \mathcal{X}; \mathbb{R}^M)$ holds for each $y \in \mathcal{Y}$, and $\partial V(\cdot, x) \in C^1(\Sigma \times \mathcal{Y}; \mathbb{R}^M)$ holds for each $x \in \mathcal{X}$. (vi) $\partial v(y, \cdot; \cdot)/\partial x \in C^1(\mathcal{X}; L^1(\Sigma))$ holds for each $y \in \mathcal{Y}$, and $\partial v(\cdot, x; \cdot)/\partial y \in C^1(\mathcal{Y}; L^1(\Sigma))$ holds for each $x \in \mathcal{X}$. (vii) X is exogenous. (viii) There exist values $p \geq 1$ and $q \geq 1$ satisfying $p^{-1} + q^{-1} = 1$ such that $\|\gamma(x, \cdot)\|_{L^p(\partial V(y, x), H^{M-1})} < \infty$ and $\|f_U\|_{L^q(\partial V(y, x), H^{M-1})} < \infty$ hold, where $\gamma(x, u) := \|\nabla_u g(x, u)\|^{-1}$.

Lemma 5 (Sasaki, 2015). If Assumption 10 is true, then the function $\mu_{y,x} : \mathcal{B}(y, x) \rightarrow \mathbb{R}$ defined by

$$\mu_{y,x}(S) := \frac{\int_S \frac{f_U(u)}{\|\nabla_u g(x, u)\|} dH^{M-1}(u)}{\int_{\partial V(y, x)} \frac{f_U(u)}{\|\nabla_u g(x, u)\|} dH^{M-1}(u)} \quad \text{for all } S \in \mathcal{B}(y, x)$$

is a probability measure on $(\partial V(y, x), \mathcal{B}(y, x))$, and the slope Q'_τ of the CQF at $\tau = F_{Y|X}(y|x)$ can be written as the weighted mean structural partial effect with respect to $\mu_{y,x}$:

$$Q'_\tau(x) = E_{\mu_{y,x}} \left[\frac{\partial g(x, U)}{\partial x} \right].$$

To make this auxiliary result applicable to our framework, we take an additional step of translating the constructed probability measure $\mu_{y,x}$ into a regular conditional probability measure on $\mathcal{B}(\mathcal{U})$ given the events in $\mathcal{B}(\mathcal{X})$. We define the function $\mu_\tau : \mathcal{B}(\mathcal{U}) \times \mathcal{X} \rightarrow [0, 1]$

by

$$\mu_\tau(S, x) = \mu_{Q_\tau(x), x}(S \cap \partial V(Q_\tau(x), x)). \quad (2.4.1)$$

By the following lemma, claiming that this μ_τ is a regular conditional probability measure, we can rewrite the conclusion of Lemma 5 formally in terms of the conditional expectation

$$Q'_\tau(x) = E_{\mu_\tau} \left[\frac{\partial g(X, U)}{\partial x} \middle| X = x \right]. \quad (2.4.2)$$

Lemma 6. *If Assumption 10 is satisfied, then μ_τ is a regular conditional probability measure on $\mathcal{B}(\mathcal{U})$ given the events in $\mathcal{B}(\mathcal{X})$.*

A proof of this lemma is found in Section B.8 in the appendix. We remark that this weighted measure μ_τ need not be the unique conditional probability measure that relates Q'_τ and $\partial g / \partial x$. For instance, Chernozhukov, Fernández-Val, Hoderlein, Holzmann, and Newey (2015) show that Q'_τ can be written as an unweighted average of $\partial g / \partial x$ under a partial monotonicity assumption in the presence of multiple unobservables. Since the measure μ_τ does not necessarily simplify to the uniform measure under the partial monotonicity assumption, weights connecting Q'_τ and $\partial g / \partial x$ need not be unique.

Thus far, we have developed a marginal probability measure λ_{x_1, x_2} on $\mathcal{B}(\mathcal{X})$ defined in (2.2.1), and a regular conditional probability measure μ_τ on $\mathcal{B}(\mathcal{U})$ given the events in $\mathcal{B}(\mathcal{X})$ defined in (2.4.1). Combining the respective mean (2.2.4) and the conditional mean (2.4.2) using the law of iterated expectations, we obtain

$$\beta(\tau) = E_{\lambda_{x_1, x_2}} \left[E_{\mu_\tau} \left[\frac{\partial g(X, U)}{\partial x} \middle| X \right] \right] = E_{\mu_\tau \times \lambda_{x_1, x_2}} \left[\frac{\partial g(X, U)}{\partial x} \right].$$

We formally state this result as the following theorem.

Theorem 5 (Weighted Average: General Structural Function). *If Assumptions 5, 6, and 10 are true, then there exist a marginal probability measure λ_{x_1, x_2} on $\mathcal{B}(\mathcal{X})$ and a regular conditional*

probability measure μ_τ on $\mathcal{B}(\mathcal{U})$ given the events in $\mathcal{B}(\mathcal{X})$ such that

$$\beta(\tau) = E_{\mu_\tau \times \lambda_{x_1, x_2}} \left[\frac{\partial g(X, U)}{\partial x} \right]$$

holds.

This result formally characterizes the slope parameter $\beta(\tau)$ of the τ -th linear QR L_τ as a weighted average of the structural partial effects $\partial g / \partial x$ with respect to the measure $\mu_\tau \times \lambda_{x_1, x_2}$. Because the weights are strictly positive on the interval $[x_1, x_2]$, the slope parameter of the linear QR represents a strict convex combination of the structural partial effects within the interval. This implies that, if $\beta(\tau)$ is positive, for example, there exists a nontrivial subpopulation in $[x_1, x_2]$ such that the structural partial effects are positive. In other words, even if the linear QR is quite different from the structural function, the linear QR can be useful for inference of the structural causal effects.

2.5 Linear Regressions and Structural Functions

While the main contribution of this paper is about quantile regressions, we show in this section that the linear regression parameter can be similarly expressed as a weighted average of the structural partial effects. Let λ_U be the probability measure of U supported on \mathcal{U} . We state the following assumption.

Assumption 11. (i) g is continuously differentiable. (ii) $g(x, \cdot) \in \mathcal{L}^1(\lambda_U)$ for every $x \in \mathcal{X}$. (iii) There exists some function $h \in \mathcal{L}^1(\lambda_U)$ such that $\left| \frac{\partial g(x, u)}{\partial x} \right| \leq h(u)$ for all $(x, u) \in \mathcal{X} \times \mathcal{U}$.

Under this assumption, the slope of the CEF equals the conditional mean of the structural partial effects, i.e.,

$$m'(x) = E_{\lambda_U} \left[\frac{\partial g(X, U)}{\partial x} \middle| X = x \right].$$

Thus, from Corollary 2, we obtain

$$\beta = E_{\lambda_{x_1, x_2}} \left[E_{\lambda_U} \left[\frac{\partial g(X, U)}{\partial x} \middle| X \right] \right] = E_{\lambda_U \times \lambda_{x_1, x_2}} \left[\frac{\partial g(X, U)}{\partial x} \right]$$

We formally state this result in the following theorem.

Theorem 6 (Weighted Average: Linear Regression and General Structural Function). *If Assumptions 7 and 11 are true, then there exists a probability measure λ_{x_1, x_2} on $\mathcal{B}(\mathcal{X})$ such that*

$$\beta = E_{\lambda_U \times \lambda_{x_1, x_2}} \left[\frac{\partial g(X, U)}{\partial x} \right]$$

holds.

This result formally characterizes the slope parameter β of the linear regression L_m as a weighted average of the structural partial effects $\partial g / \partial x$ with respect to the measure $\lambda_U \times \lambda_{x_1, x_2}$. Because the weights are strictly positive on the interval $[x_1, x_2]$, the slope parameter of the linear regression represents a strict convex combination of the structural partial effects within the interval. Therefore, even if the linear regression is quite different from the structural function, it can be useful for inference of the structural causal effects.

2.6 Conclusion

The slope parameter of the linear QR can be written as weighted averages of the local slopes of the CQF among a nontrivial subpopulation of individuals. Likewise, the slope parameter of the linear regression can be written as weighted averages of the local slopes of the CEF among a nontrivial subpopulation of individuals. These results follow from the property that the linear and nonlinear functions related by weighted least squares must have at least two intersection points. We present an intuitive illustration as well as a formal derivation of this result.

When the structural function g is monotone with respect to a scalar latent variable U , the structural function can be represented by the CQF Q_τ . Therefore, our first result directly implies that the slope parameter $\beta(\tau)$ of the τ -th linear QR measures a weighted average of the structural partial effects $\partial g / \partial x$. Furthermore, even if the structural function g fails to exhibit monotonicity, we establish that the slope parameter $\beta(\tau)$ still measures a weighted average of the structural partial effects $\partial g / \partial x$. We also obtain a similar result for the linear regression.

From these findings, we conclude that the linear regression and the linear QR can be used for causal inference even if they misspecify the true structure both in terms of nonlinearity and non-monotonicity.

Bibliography

- [1] Angrist, J., V. Chernozhukov and I. Fernández-Val (2006) “Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure,” *Econometrica*, 74 (2), 539–563.
- [2] Angrist, J.D. and G.W. Imbens (1995) “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” *Journal of the American Statistical Association*, 90 (430), 431–442.
- [3] Angrist, J.D. and A.B. Krueger (1999) “Empirical Strategies in Labor Economics,” in O.C. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3, Part A, 1277–1366.
- [4] Chamberlain, G. (1984) “Panel Data,” in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, 1247–1318.
- [5] Chernozhukov, V., I. Fernández-Val and A. Galichon (2010) “Quantile and Probability Curves Without Crossing,” *Econometrica*, 74 (2), 539–563.
- [6] Chernozhukov, V., I. Fernández-Val, S. Hoderlein, H. Holzmann and W. Newey (2015) “Nonparametric Identification in Panels Using Quantiles,” *Journal of Econometrics*, 188 (2), 378–392.
- [7] Chernozhukov, V. and C. Hansen (2005) “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73 (1), 245–261.

- [8] Chernozhukov, V. and C. Hansen (2013) "Quantile Models with Endogeneity," *Annual Review of Economics*, 5 (1), 57–81.
- [9] Goldberger, A.S. (1991) *A Course in Econometrics*. Harvard University Press: Cambridge.
- [10] Imbens, G.W. and W.K. Newey (2009) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77 (5), 1481–1512.
- [11] Koenker, R. (2005) "Quantile Regression," Cambridge University Press: Cambridge.
- [12] Koenker, R. and G. Bassett, Jr. (1978) "Regression Quantiles," *Econometrica*, Vol. 46, No. 1, pp. 33–50.
- [13] Koenker, R. and K.F. Hallock (2001) "Quantile Regression," *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143–156.
- [14] Lee, Y.-Y. (2013) "Interpretation and Semiparametric Efficiency in Quantile Regression under Misspecification," Working Paper, University of Wisconsin-Madison.
- [15] Sasaki, Y. (2015) "What Do Quantile Regressions Identify for General Structural Functions?" *Econometric Theory*, Vol. 31, No. 5, pp. 1102–1116.
- [16] White, H. (1980) "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21 (1), 149–170.
- [17] Yitzhaki, S. (1996) "On Using Linear Regressions in Welfare Economics," *Journal of Business and Economic Statistics*, 14 (4), 478–486.

Chapter 3

IV Quantile Regression under Misspecification

3.1 Introduction

The linear quantile regression has been an important tool to model unobserved heterogeneity. While the linear quantile regression is an attractive tool to capture the quantile specific effects, there are two major obstacles. One is endogeneity and the other is misspecification. There are several interpretation of the linear quantile regression when linear model of the conditional quantile function is misspecified. Angrist, Chernozhukov and Fernandez-Val (2006) show that the parameters of the linear quantile regression minimize the weighted average of squared distances between the linear quantile regression. This result parallels the feature of the ordinary least squares (OLS) estimates that minimize the average squares distance between the linear regression and the true conditional regression. Kato and Sasaki (2015) extends their result and shows that the slope parameter of the linear quantile regression represents a weighted average of the slopes of the conditional quantile function. They also show that slope parameter of the linear quantile regression represents a weighted average of the slopes of the general structural function. Lee (2014) also gives another fit of measure for the linear quantile regression parameter. However these interpretation assumes that

the variable of interest is exogenous. In many cases in economics, variables of interests are endogeneous making the quantile regression method inappropriate to recover the causal effects of these variables on the quantiles of the outcome. Chernozhukov and Hansen (2005) develop a model of quantile regressions in the presence of endogeneity and obtain conditions for identification of the quantile regression without imposing any functional assumptions. Chernozhukov and Hansen (2006) gives an estimation strategy of the quantile regression in the presence of endogeneity.

In this paper, I analyze misspecification problem in the presence of endogeneity. We study the relation between the linear quantile regression and the true quantile regression when the variable of interest is endogenous. First, I show that when the linear model of the quantile regression is misspecified, the slope parameter of the linear quantile regression minimizes a weighted mean squared error loss function for specification error. This result is analogous to the result given in Angrist, Chernozhukov and Fernandez-Val (2006) which shows that parameters of the linear quantile regression minimize the weighted average of squared distances between the linear quantile regression when the variable of interest is exogenous. Extending this result, I show that the slope of the linear quantile regression represents a weighted average of the slope of the true quantile regression. This result coincides with the result given in Kato and Sasaki (2015) which shows that the slope parameter of the linear quantile regression represents a weighted average of the slopes of the true conditional quantile function when the variable of interest is exogenous.

3.2 Linear Quantile regression and Quantile regression

We first introduce some basic definitions and notations. Let Y be the outcome variable, X denote the treatment, which is possibly endogenous, and Z denote the instrumental variable

supported on $\mathcal{X} \in \mathbb{R}$, $\mathcal{Y} \in \mathbb{R}$ and $\mathcal{Z} \in \mathbb{R}$ respectively. For a given $\tau \in (0, 1)$, the τ -th quantile regression Q_τ is implicitly defined as

$$E[(1\{Y < Q_\tau(X)\} - \tau)|Z] = 0 \text{ a.s.}$$

The τ -th linear quantile regression L_τ is given by minimizing the GMM criterion function. In other words, $L_\tau(X) = \alpha(\tau) + \beta(\tau) \cdot X$ where the parameters $(\alpha(\tau), \beta(\tau))$ is given by

$$(\alpha(\tau), \beta(\tau)) = \arg \min_{(\alpha, \beta)} E[(1\{Y < Q_\tau(X)\} - \tau) \cdot Z]$$

In general, $L_\tau \neq Q_\tau$. Also define a short hand notation

$$\Delta_\tau(x, \alpha(\tau), \beta(\tau)) : = \alpha(\tau) + \beta(\tau)x - Q_\tau(x) = L_\tau(x) - Q_\tau(x).$$

$$\epsilon_t : = Y - Q_\tau(X)$$

I make a connection between Q_τ and L_τ in the following section.

3.2.1 Weighted Mean Squared Error Minimization

The main result is that under regularity assumptions the slope $\beta(\tau)$ of the τ -th linear quantile regression L_τ can be explicitly written as a weighted average of the slopes Q'_τ of the quantile regression Q_τ among a non-trivial subpopulation of individuals even when the treatment is endogenous. To derive this conclusion, I start by showing that the slope of the linear quantile regression minimizes the weighted means square approximation error between the true quantile regression under the following assumption.

Assumption 12. (i) the conditional density $f_Y(y|XZ)$ exists and is bounded a.s, (ii) $E[Y]$, $E[Q_\tau(X)]$ and $E\|X\|^2$ are finite, (iii) $\alpha(\tau), \beta(\tau)$ uniquely solves $\min_{\alpha, \beta} (E[(1\{Y < \alpha + \beta X\} - \tau) \cdot Z])^2$ (iv) $Q_\tau(x)$ is continuous

(v) $Q_\tau(X)$ is identified i.e. $E[(q(X) - Q_\tau(X)) \cdot \int_0^1 f_Y(u \cdot (q(x) - Q_\tau(x)) | X, Z) du] = 0$ a.s implies that $q(x) = Q_\tau(x)$ a.s..

Many parts of Assumption 12 are inherited from the assumptions stated in Theorem 2 of Angrist, Chernozhukov and Fernández-Val (2006).

Theorem 7. *If Assumption 12 is true, then the slope of the linear quantile regression $\beta(\tau)$ uniquely solves*

$$\min_{\beta} E[\tilde{\omega}_\tau(X, Z, \alpha(\tau), \beta(\tau)) \Delta_\tau^2(X, \alpha(\tau), \beta)]$$

where

$$\tilde{\omega}_\tau(X, Z, \alpha(\tau), \beta(\tau)) = \frac{Z}{X} \cdot \int_0^1 f_Y(u(\alpha(\tau) + \beta(\tau)x) + (1 - u)Q_{Y|X}(\tau|x) | XZ) du$$

The proof of this Theorem is provided in the appendix. Theorem 1 states that the slope of the linear quantile regression minimizes the weighted means square approximation error between the true quantile regression. This result is similar to the result given in Angrist, Chernozhukov and Fernandez-Val (2006) which analyzes the misspecification problem under exogeneity. Theorem 1 implies that even in the presence of endogeneity, linear quantile regression would have a similar approximation property.

3.2.2 Weighted Average Interpretation

By extending Theorem 7, we obtain an auxiliary lemma that for any $\tau \in (0, 1)$ the linear quantile regression L_τ intersects with the quantile regression Q_τ at least at two points of x in the support \mathcal{X} of X under the following additional assumption.

Assumption 13. (i) $Q_\tau(0) = \alpha(\tau)$ (ii) $\mathcal{X} \subset \mathbb{R}_+$ (iii) $0 \in \mathcal{X}$ (iv) \mathcal{X} is convex. (v) $\mathcal{Z} \subset \mathbb{R}_+$

Assumption 13(i) normalizes the true quantile regression Q_τ to ensure that Q_τ and L_τ intersects at least at one point.

Lemma 7. *If Assumption 12 and Assumption 13 are true, L_τ and Q_τ intersect at least at two points in \mathcal{X} .*

The proof of this lemma is shown in the appendix. It is proved analogously to the logic used in the proof of Lemma 7 of Kato and Sasaki (2015) that they developed in the case of exogenous X . To claim that the slope $\beta(\tau)$ of the τ -th linear quantile regression L_τ is a weighted average of the slopes Q'_τ of the quantile regression, we now invoke an additional assumption that ensures that this derivative Q'_τ of interest exists.

Assumption 14. *Q_τ is continuously differentiable.*

Applying Lemma 7 and the Fundamental Theorem of Calculus, we now obtain the following result.

Theorem 8. *If Assumptions 1, 13 and 14 are true, then there exist $x_1 \in \mathcal{X}$ with such that*

$$\beta(\tau) = E [w_{x_1}(X) \cdot Q'_\tau(X) | 0 \leq X \leq x_1]$$

holds, where the weight function w_{x_1} is given by

$$w_{x_1}(x) := \frac{\frac{1}{f_X(x)}}{\int_0^{x_1} \frac{1}{f_X(\xi)} dF_X(\xi)} \quad \text{for all } x \in \mathcal{X}.$$

The proof is a straight forward application of the Fundamental Theorem of Calculus. This theorem implies that the slope of the linear quantile regression represents the weighted average of the slope of the true quantile regression. Note that the weight function w_{x_1} is strictly positive in $[0, x_1]$. In other words, the slope of the linear quantile regression represents a strict convex combination of the slopes of the true quantile regression within an interval $[0, x_1] \in \mathcal{X}$. To formalize the weighted-average interpretation, we define the function $\lambda_{x_1, x_2} : \mathcal{B}(\mathcal{X}) \rightarrow \mathbb{R}$ by

$$\lambda_{x_1, x_2}(T) = \int_{T \cap [x_1, x_2]} w_{x_1, x_2}(x) dF_X(x) \quad \text{for each } T \in \mathcal{B}(\mathcal{X}). \quad (3.2.1)$$

Note that this λ_{x_1, x_2} can be shown to be a probability measure. With this notation, the conclusion of Theorem 2 can be restated as

$$\beta(\tau) = E_{\lambda_{x_1, x_2}} [Q'_\tau(X)] . \quad (3.2.2)$$

In other words, the slope $\beta(\tau)$ of the τ -th linear quantile regression is the mean of the slopes Q'_τ of the quantile regression with respect to the probability measure λ_{x_1, x_2} .

3.2.3 Discrete Case

The main result (Theorem 8) was obtained assuming that the regressor X is continuously distributed. In the current sub-section, we show that a similar idea applies to the case of a discrete regressor, which is very common in economic applications. As we move from a continuous distribution to a discrete distribution, we also change the object of interest from derivatives to differences. The following short-hand notations are introduced.

$$\beta(\tau) = \Delta L_\tau(x) \quad : \quad = L_\tau(x') - L_\tau(x)$$

$$\Delta Q_\tau(x) \quad : \quad = Q_\tau(x') - Q_\tau(x)$$

The goal is to show that the slope $\beta(\tau)$ of the linear quantile regression L_τ can be expressed as a weighted average of the differences ΔQ_τ of the conditional quantile function. To this goal, we use the following assumption.

Assumption 15. (i) $Q_\tau(0) = \alpha(\tau)$ (ii) The conditional density $f_{Y|X}(\cdot | x)$ exists and is bounded for each $x \in \mathcal{X}$. (iii) $E[Y], E[Q_\tau(X)]$ and EX^2 are finite. (iv) $(\alpha(\tau), \beta(\tau))$ uniquely solves $\min_{(\alpha, \beta)} E[\rho_\tau(Y - \alpha - \beta X)]$. (v) $\mathcal{X} \subset \mathbb{Z}$ and $|\mathcal{X}| \geq 2$. (vi) $Q_\tau(X)$ is identified i.e. $E[(q(X) - Q_\tau(X)) \cdot \int_0^1 f_Y(u \cdot (q(x) - Q_\tau(x)) | X, Z) du] = 0$ a.s

implies that $q(x) = Q_\tau(x)$ a.s.. (vii) $0 \in \mathcal{X}$

The essence of this assumption inherits from that of Assumption 12 and 13, that we used for the case of continuous X . Compared to that assumption, the current assumption drops the continuity of Q_τ and the continuous distribution of X . Instead, it adds the restriction that \mathcal{X} is integer-valued although it can be relaxed to arbitrary discrete supports with isolated points. Again I have to include the normalization assumption as I did in the continuous case. Under this assumption, we obtain the following auxiliary lemma.

Lemma 8. *If Assumption 15 is true, then there exist two distinct points $x_1, x_2 \in \mathcal{X}$ such that $(\Delta L_\tau(x_1) - \Delta Q_\tau(x_1)) \cdot (\Delta L_\tau(x_2) - \Delta Q_\tau(x_2)) \leq 0$ holds.*

A proof is found in Section C.4 in the appendix. It is proved analogously to the logic used in the proof of Lemma 7, that we developed for the case of continuous X . This lemma claims that there is some point $x_1 \in \mathcal{X}$ at which $\beta(\tau) = \Delta L_\tau(x_1) \geq \Delta Q_\tau(x_1)$ is true and another point $x_2 \in \mathcal{X}$ at which $\beta(\tau) = \Delta L_\tau(x_2) \leq \Delta Q_\tau(x_2)$ is true. In other words, the slope parameter $\beta(\tau)$ of the linear quantile regression is bounded from both above and below by the differences ΔQ_τ of the quantile function. This property immediately implies that $\beta(\tau)$ can be written as a convex combination of $\{\Delta Q_\tau(x) \mid x \in \mathcal{X}\}$, as formally stated in the following theorem.

Theorem 9. *If Assumption 15 is true, then*

$$\beta(\tau) = E[w(X) \cdot \Delta Q_\tau(X)]$$

holds for a non-negative weight function w such that $E[w(X)] = 1$.

A proof is provided in Section C.5 in the appendix. While our proof constructs a particular weight function w to establish the equality, such a weight function need not be unique.

3.3 Conclusion

I have shown that the slope of the linear quantile regression gives a weighted mean square approximation to the true quantile regression even in the presence of endogeneity. Extending this result, I also show that the slope of the linear quantile regression represents a weighted average of the local slope of the true quantile regression among a nontrivial subpopulation of individuals even when the variable of interest is endogenous. These results follow from the property that the linear and nonlinear functions related by weighted least squares must have at least two intersection points. These findings coincide with the result given by Angrist, Chernozhukov and Fernandez-Val (2006) and Kato and Sasaki (2015) which analyzes the misspecification problem of quantile regression under exogeneity. This paper shows that similar interpretation holds even in the presence of endogeneity.

Bibliography

- [1] Angrist, Joshua, Victor Chernozhukov, and Ivan Fernandez-Val. "Quantile regression under misspecification, with an application to the US wage structure." *Econometrica* 74.2 (2006): 539-563.
- [2] Angrist, J.D. and G.W. Imbens (1995) "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90 (430), 431–442.
- [3] Chernozhukov, Victor, and Christian Hansen. "An IV model of quantile treatment effects." *Econometrica* (2005): 245-261.
- [4] Chernozhukov, Victor, and Christian Hansen. "Instrumental quantile regression inference for structural and treatment effect models." *Journal of Econometrics* 132.2 (2006): 491-525.
- [5] Chernozhukov, Victor, and Christian Hansen. "Quantile models with endogeneity." *arXiv preprint arXiv:1303.7050* (2013).
- [6] Kato, Ryutah, Yuya Sasaki, and Johns Hopkins Johns Hopkins. "On Using Linear Quantile Regressions for Causal Inference." (2015).
- [7] Koenker, Roger, and Gilbert Bassett Jr. "Regression quantiles." *Econometrica: journal of the Econometric Society* (1978): 33-50.

- [8] Lee, Ying-Ying. Interpretation and Semiparametric Efficiency in Quantile Regression under Misspecification. Working paper, <https://sites.google.com/site/yyleelilian>, 2013.

Appendix A

Mathematical Appendix of Chapter 1

A.1 Proof or Proposition 1

Proof. Recall that the value of the firm M_{t+1} is

$$M_{t+1} = \max_{\{d_{t+s}\}_{s=1}^{\infty}} E_{t+1} \left[\sum_{s=1}^{\infty} \beta^s \cdot \pi_{t+s} - d_{t+s} \cdot C_f \right] \quad (\text{A.1.1})$$

Let $\epsilon_t := \theta - \mu_{t-1} + \eta_t$. Note that $E_t[\epsilon_t] = 0$. Then we can write π_{t+1} as

$$\pi_{t+1} = \phi \cdot \pi_t + \mu_t + \epsilon_{t+1} \quad (\text{A.1.2})$$

Then we have

$$\pi_{t+s} = \phi^s \cdot \pi_t + \sum_{k=0}^{s-1} \mu_{t+k} \cdot \phi^{s-1-k} + \sum_{k=0}^{s-1} \epsilon_{t+k} \cdot \phi^{s-1-k} \quad (\text{A.1.3})$$

Since $E_t[\epsilon_{t+k}] = E_t[E_{t+k}[\epsilon_{t+k}]] = 0$, we have

$$E_{t+1}[\pi_{t+s}] = \phi^s \cdot \pi_{t+s} + E_{t+1} \left[\sum_{k=0}^{s-1} \mu_{t+k} \cdot \phi^{s-1-k} \right] \quad (\text{A.1.4})$$

Next we have

$$E_{t+1}[\sum_{s=1}^{\infty} \beta^{s-1} \cdot \pi_{t+s}] = \frac{\phi}{1 - \beta \cdot \phi} \cdot \pi_t + \frac{1}{1 - \beta \cdot \phi} \cdot \sum_{k=0}^{s-1} \beta^{s-1} \cdot E_{t+1}[\mu_{t+k}] \quad (\text{A.1.5})$$

Therefore the firm value can be written as

$$M_{t+1} = \frac{\phi}{1 - \beta \cdot \phi} \cdot \pi_t + V(\mu_t, \tau_t) \quad (\text{A.1.6})$$

where

$$V(\mu_t, \tau_t) = \max_{d_t} \left\{ \frac{1}{1 - \beta \cdot \phi} \cdot \mu_t - d_t \cdot C_f + \beta \cdot E[V(\mu_t, \tau_t)] \right\} \quad (\text{A.1.7})$$

If $d_t = 1$ (the firm fires the CEO), then the firm hires a new CEO and pays the firing cost

$$V_f(\mu_t, \tau_t) = V(\mu_0, 0) - C_f \quad (\text{A.1.8})$$

If $d_t = 0$ (the firm keeps the CEO), then

$$V_k(\mu_t, \tau_t) = \frac{1}{1 - \beta \cdot \phi} \cdot \mu_t + \beta \cdot E[V(\mu_{t+1}, \tau_{t+1})] \quad (\text{A.1.9})$$

The CEO retires after T years at the office and the firm hires a new CEO. Hence if $\tau_t = T$ we have

$$V_k(\mu_t, T) = V(\mu_0, 0) \quad (\text{A.1.10})$$

The firm chooses d_t according to

$$V(\mu_t, \tau_t) = \max\{V_k(\mu_t, \tau_t), V_f(\mu_t, \tau_t)\} \quad (\text{A.1.11})$$

From equation(1.2.4), if $d_t = 0$, μ_t and τ_t follows

$$\mu_{t+1} = \mu_t + \frac{\sigma^2(\tau_t)}{\sigma_\eta^2} \cdot \epsilon_{t+1} \quad (\text{A.1.12})$$

$$\tau_{t+1} = \begin{cases} \tau_t + 1 & \text{if } (0 \leq \tau_t \leq T - 1) \\ 1 & \text{if } (\tau_t = T) \end{cases} \quad (\text{A.1.13})$$

where

$$\sigma^2(\tau_t) = \sigma_0^2 \cdot \left(1 + \tau_t \cdot \frac{\sigma_0^2}{\sigma_\eta^2} \right)^{-1} \quad (\text{A.1.14})$$

Substitute equation (A.1.12)-(A.13) to equation (A.1.9) completes the proof.

□

Appendix B

Mathematical Appendix of Chapter 2

B.1 Proof of Lemma 1

Proof. By way of contradiction, suppose that M and L intersect at most at one point of x .

Define the short-hand notation

$$\Delta(x, \alpha, \beta) := \alpha + \beta x - M(x) = L(x) - M(x).$$

Under the current assumption, we have at most one zero for $\Delta(\cdot, \alpha, \beta)$ in \mathcal{X} . By Assumption 3 (iv)-(v), it follows that either one of the following three cases is true.

- (I) There exists a point $x^* \in \mathcal{X}$ such that $\Delta(x, \alpha, \beta) \cdot \Delta(x', \alpha, \beta) < 0$ for all $x \in \mathcal{X} \cap (x^*, \infty)$ and for all $x' \in \mathcal{X} \cap (-\infty, x^*)$.
- (II) There exists a point $x^* \in \mathcal{X}$ such that $\Delta(x^*, \alpha, \beta) = 0$, but $\Delta(x, \alpha, \beta) \cdot \Delta(x', \alpha, \beta) > 0$ for all $x, x' \in \mathcal{X}$ such that $x \neq x^*$ and $x' \neq x^*$.
- (III) $\Delta(x, \alpha, \beta) \cdot \Delta(x', \alpha, \beta) > 0$ for all $x, x' \in \mathcal{X}$.

We claim below that each of the cases (I)–(III) contradicts Assumption 3 (ii) that (α, β) uniquely solves $\min_{(a,b)} E[\omega(X) \cdot (M(X) - a - bX)^2]$.

First, consider case (I). Without loss of generality, we normalize the location to $x^* = 0$ and assume $\Delta(x, \alpha, \beta) < 0$ for all $x > 0$. For each $\epsilon > 0$, define the set

$$A(\epsilon) = \{x \in \mathbb{R}_{++} \cap \mathcal{X} \mid \Delta(x, \alpha, \beta + \epsilon) \leq 0\} \cup \{x \in \mathbb{R}_{--} \cap \mathcal{X} \mid \Delta(x, \alpha, \beta + \epsilon) \geq 0\}.$$

We let B^1 and B^2 denote arbitrary compact intervals contained in $\mathcal{X} \cap \mathbb{R}_{--}$ and $\mathcal{X} \cap \mathbb{R}_{++}$, respectively. Note that $\min_{x \in B^1 \cup B^2} \left\{ \frac{\Delta(x, \alpha, \beta)}{-x} \right\}$ exists due to the compactness of $B^1 \cup B^2$ and the continuity of $\frac{\Delta(x, \alpha, \beta)}{-x}$ with respect to x on $B^1 \cup B^2$. Thus, if we choose $\epsilon := \min_{x \in B^1 \cup B^2} \left\{ \frac{\Delta(x, \alpha, \beta)}{-x} \right\}$, then $\Delta(x, \alpha, \beta + \epsilon) \geq 0$ for all $x \in B^1$ and $\Delta(x, \alpha, \beta + \epsilon) \leq 0$ for all $x \in B^2$. Hence, for any compact intervals $B^1 \subset \mathcal{X} \cap \mathbb{R}_{--}$ and $B^2 \subset \mathcal{X} \cap \mathbb{R}_{++}$, there exists $\epsilon > 0$ such that $(B^1 \cup B^2) \subset A(\epsilon)$. Furthermore, note that $\epsilon > 0$ is true. Now, observe from the definition of $A(\epsilon)$ that the inequality

$$\begin{aligned} & E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha, \beta + \epsilon) \cdot \omega(x)] \\ &= \int_{\mathcal{X}} (-2\Delta(x, \alpha, \beta)\epsilon x - (\epsilon x)^2) \cdot \omega(x) dF_X(x) \\ &\geq \int_{A(\epsilon)} (\epsilon x)^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon)} (\epsilon x)^2 \cdot \omega(x) dF_X(x) \end{aligned} \quad (\text{B.1.1})$$

holds for any $\epsilon > 0$. Let $B_n := \mathcal{X} \cap \left(\left[-n, -\frac{1}{n} \right] \cup \left[\frac{1}{n}, n \right] \right)$ for each integer $n > 1$. Then for each $n > 1$, the above argument implies that there exists $\epsilon_n > 0$ such that $B_n \subset A(\epsilon_n)$. But then, for each $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} & \int_{B_n} x^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus B_n} x^2 \cdot \omega(x) dF_X(x) \\ &\leq \int_{A(\epsilon_n)} x^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon_n)} x^2 \cdot \omega(x) dF_X(x). \end{aligned} \quad (\text{B.1.2})$$

By Assumption 3 (i) and (iii) and the Monotone Convergence Theorem,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \int_{\mathcal{X} \setminus B_n} x^2 \cdot \omega(x) dF_X(x) \\
&= \lim_{n \rightarrow \infty} \int_{\mathcal{X}} x^2 \cdot \omega(x) \cdot f_X(x) \cdot \mathbb{1}\{x \in \mathcal{X} \setminus B_n\} dx \\
&= \int_{\mathcal{X}} \lim_{n \rightarrow \infty} x^2 \cdot \omega(x) \cdot f_X(x) \cdot \mathbb{1}\{x \in \mathcal{X} \setminus B_n\} dx = 0
\end{aligned} \tag{B.1.3}$$

is true. Likewise, Assumption 3 (i) and (iii) and the Monotone Convergence Theorem yield

$$\lim_{n \rightarrow \infty} \int_{B_n} x^2 \cdot \omega(x) dF_X(x) = \int_{\mathcal{X}} x^2 \cdot \omega(x) \cdot f_X(x) dx =: c. \tag{B.1.4}$$

Note that $c \geq 0$, and it holds with equality only if $\omega(x) \cdot f_X(x) = 0$ almost everywhere on \mathcal{X} . But it is not true that $\omega(x) \cdot f_X(x) = 0$ almost everywhere on \mathcal{X} , from Assumption 3 (ii) that (α, β) uniquely minimizes the weighted mean squared distance. Thus, it follows from (C.2.1)–(C.2.4) that

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha, \beta + \epsilon_n) \cdot \omega(x)]}{\epsilon_n^2} \\
& \geq c > 0
\end{aligned}$$

is true. But then, there exists $n^* > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha, \beta + \epsilon_{n^*}) \cdot \omega(x)]}{\epsilon_{n^*}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] > E[(\Delta^2(X, \alpha, \beta + \epsilon_{n^*}) \cdot \omega(x)],$$

and it contradicts Assumption 3 (ii), that (α, β) uniquely minimizes the weighted mean squared distance.

Next, consider case (II). Without loss of generality, we normalize the location to $x^* = 0$ and assume $\Delta(x, \alpha, \beta) > 0$ for all $x \neq 0$. For each $\epsilon > 0$, define the set

$$A'(\epsilon) = \{x \in \mathcal{X} \mid \Delta(x, \alpha - \epsilon, \beta) \geq 0\}.$$

Let B^1 and B^2 denote arbitrary compact intervals contained in $\mathcal{X} \cap \mathbb{R}_{--}$ and $\mathcal{X} \cap \mathbb{R}_{++}$, respectively. Note that $\min_{x \in B^1 \cup B^2} \{\Delta(x, \alpha, \beta)\}$ exists due to the compactness of $B^1 \cup B^2$ and the continuity of $\Delta(x, \alpha, \beta)$ with respect to x on $B^1 \cup B^2$. If we choose $\epsilon := \min_{x \in B^1 \cup B^2} \{\Delta(x, \alpha, \beta)\}$, then $\Delta(x, \alpha - \epsilon, \beta) \geq 0$ for all $x \in B^1 \cup B^2$. Hence, for any compact intervals $B^1 \subset \mathcal{X} \cap \mathbb{R}_{--}$ and $B^2 \subset \mathcal{X} \cap \mathbb{R}_{++}$, there exists $\epsilon > 0$ such that $(B^1 \cup B^2) \subset A'(\epsilon)$. Furthermore, note that

$\epsilon > 0$ is true. It follows that the inequality

$$\begin{aligned}
& E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[\Delta^2(X, \alpha - \epsilon, \beta) \cdot \omega(x)] \\
&= \int_{\mathcal{X}} (2\Delta(x, \alpha, \beta)\epsilon - \epsilon^2) \cdot \omega(x) dF_X(x) \\
&\geq \int_{A'(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A'(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) \tag{B.1.5}
\end{aligned}$$

holds for any $\epsilon > 0$. By the same argument as the one used in case (I), for each integer $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned}
& \int_{B_n} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) \\
&\leq \int_{A'(\epsilon_n)} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A'(\epsilon_n)} \omega(x) dF_X(x), \tag{B.1.6}
\end{aligned}$$

where

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) = 0 \quad \text{and} \tag{B.1.7}$$

$$\lim_{n \rightarrow \infty} \int_{B_n} \omega(x) dF_X(x) = \int_{\mathcal{X}} \omega(x) \cdot f_X(x) dx =: c' > 0. \tag{B.1.8}$$

Thus, it follows from (B.1.5)–(B.1.8) that

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha - \epsilon_n, \beta) \cdot \omega(x)]}{\epsilon_n^2} \\
&\geq c' > 0
\end{aligned}$$

is true. But then, there exists $n^{**} > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha - \epsilon_{n^{**}}, \beta) \cdot \omega(x)]}{\epsilon_{n^{**}}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] > E[(\Delta^2(X, \alpha - \epsilon_{n^{**}}, \beta) \cdot \omega(x)],$$

and it contradicts Assumption 3 (ii), that (α, β) uniquely minimizes the weighted mean squared distance.

Lastly, consider case (III). Without loss of generality, assume that $\Delta(x, \alpha, \beta) > 0$ for all $x \in \mathcal{X}$. As in case (II), if B^1 and B^2 are arbitrary compact intervals contained in $\mathcal{X} \cap \mathbb{R}_{--}$ and

$\mathcal{X} \cap \mathbb{R}_{++}$, respectively, then there exists $\epsilon > 0$ such that $(B^1 \cup B^2) \subset A'(\epsilon)$. Thus,

$$\begin{aligned} & E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta(X, \alpha - \epsilon, \beta))^2 \cdot \omega(x)] \\ & \geq \int_{A'(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A'(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) \end{aligned} \quad (\text{B.1.9})$$

holds for any $\epsilon > 0$ similarly to case (II). By the same argument as the one used in case (I), for each integer $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} & \int_{B_n} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) \\ & \leq \int_{A'(\epsilon_n)} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A'(\epsilon_n)} \omega(x) dF_X(x), \end{aligned} \quad (\text{B.1.10})$$

where

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) = 0 \quad \text{and} \quad (\text{B.1.11})$$

$$\lim_{n \rightarrow \infty} \int_{B_n} \omega(x) dF_X(x) = \int_{\mathcal{X}} \omega(x) \cdot f_X(x) dx =: c' > 0. \quad (\text{B.1.12})$$

Thus, it follows from (B.1.9)–(B.1.12) that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha - \epsilon_n, \beta) \cdot \omega(x)]}{\epsilon_n^2} \\ & \geq c' > 0 \end{aligned}$$

is true, similarly to case (II). But then, there exists $n^{**} > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] - E[(\Delta^2(X, \alpha - \epsilon_{n^{**}}, \beta) \cdot \omega(x)]}{\epsilon_{n^{**}}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(x)] > E[(\Delta^2(X, \alpha - \epsilon_{n^{**}}, \beta) \cdot \omega(x)],$$

and it contradicts Assumption 3 (ii), that (α, β) uniquely minimizes the weighted mean squared distance. \square

B.2 Proof of Theorem 1

Proof. From Lemma 1, Assumption 3 guarantees the existence of two points $x_1, x_2 \in \mathcal{X}$ such that $L(x_1) = M(x_1)$ and $L(x_2) = M(x_2)$ are both true. Without loss of generality, let $x_1 < x_2$.

By Assumption 4, we have $M(x_2) - M(x_1) = \int_{x_1}^{x_2} M'(\xi) d\xi$ by the Fundamental Theorem of Calculus. Similarly, we have $L(x_2) - L(x_1) = \int_{x_1}^{x_2} L'_\tau(\xi) d\xi = \beta \cdot (x_2 - x_1)$. Combining all these equalities together yields

$$\int_{x_1}^{x_2} M'(\xi) d\xi = \beta \cdot (x_2 - x_1).$$

By Assumption 3 (v), $f_X > 0$ almost everywhere on $[x_1, x_2]$. Thus, we can write

$$\beta = \frac{\int_{x_1}^{x_2} M'(\xi) d\xi}{\int_{x_1}^{x_2} d\xi} = \frac{\int_{x_1}^{x_2} \frac{M'(\xi)}{f_X(\xi)} dF_X(\xi)}{\int_{x_1}^{x_2} \frac{1}{f_X(\xi)} dF_X(\xi)} = \mathbb{E} [w_{x_1, x_2}(X) \cdot M'(X) | x_1 \leq X \leq x_2].$$

□

B.3 Proof of Lemma 2

Proof. By way of contradiction, suppose that $\beta - \Delta M(x) < 0$ is true for all $x \in \mathcal{X}$. We consider the following two cases.

(I) There exist x, x' such that $L(x) - M(x) < 0$ and $L(x') - M(x') > 0$.

(II) $L(x) - M(x) \geq 0$ for all $x \in \mathcal{X}$.

We derive a contradiction under each of these two cases to complete a proof. Specifically, we show below that each of the cases (I) and (II) contradicts Assumption 8 (ii) that (α, β) uniquely solves $\min_{(a,b)} \mathbb{E}[\omega(X) \cdot (M(X) - a - bX)^2]$.

First, we consider case (I). Since $\beta - \Delta M(x) < 0$ for all $x \in \mathcal{X}$, if $\Delta(x, \alpha, \beta) < 0$ for some $x \in \mathcal{X}$, then $\Delta(\tilde{x}, \alpha, \beta) < 0$ for all $\tilde{x} > x$. Also, if $\Delta(x', \alpha, \beta) > 0$ for some $x' \in \mathcal{X}$, then $\Delta(\tilde{x}, \alpha, \beta) > 0$ for all $\tilde{x} < x'$. Hence, there exists a unique $x^* \in \mathcal{X}$ such that $\Delta(x, \alpha, \beta) \geq 0$ for all $x \leq x^*$ and $\Delta(x, \alpha, \beta) < 0$ for all $x > x^*$. Without loss of generality, we normalize the

location to $x^* = 0$. For each $\epsilon > 0$, define the set

$$A(\epsilon) = \{x \in \mathcal{X} \cap \mathbb{Z}_+ \mid \Delta(x, \alpha, \beta + \epsilon) \leq 0\} \cup \{x \in \mathcal{X} \cap \mathbb{Z}_- \mid \Delta(x, \alpha, \beta + \epsilon) \geq 0\}.$$

We let B denote an arbitrary nonempty finite subset of \mathcal{X} .

If we choose $\epsilon := \min_{x \in B} \left\{ \frac{\Delta(x, \alpha, \beta)}{-x} \right\}$, then $\Delta(x, \alpha, \beta + \epsilon) \geq 0$ for all $x \in B \cap \mathcal{X}_+$ and $\Delta(x, \alpha, \beta + \epsilon) \leq 0$ for all $x \in B \cap \mathcal{X}_-$. Hence, for any finite subset B , there exists $\epsilon > 0$ such that $B \subset A(\epsilon)$. Furthermore, note that $\epsilon > 0$ is true. Now, observe from the definition of $A(\epsilon)$ that the inequality

$$\begin{aligned} & E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha, \beta + \epsilon) \cdot \omega(x))] \\ &= \sum_{x \in \mathcal{X}} (-2\Delta(x, \alpha, \beta)\epsilon x - (\epsilon x)^2) \cdot \omega(x) \cdot p(x) \\ &\geq \sum_{x \in A(\epsilon)} (\epsilon x)^2 \cdot \omega(x) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus A(\epsilon)} (\epsilon x)^2 \cdot \omega(x) \cdot p(x) \end{aligned} \quad (\text{B.3.1})$$

holds for any $\epsilon > 0$. Let $B_n := \{x \in \mathcal{X} \mid -n \leq x \leq n\}$ for each integer $n > 1$. Then for each, $n > 1$, the preceding argument implies that there exists $\epsilon_n > 0$ such that $B_n \subset A(\epsilon_n)$. But then, for each $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} & \sum_{x \in B_n} x^2 \cdot \omega(x) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus B_n} x^2 \cdot \omega(x) \cdot p(x) \\ &\leq \sum_{x \in A(\epsilon_n)} x^2 \cdot \omega(x) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus A(\epsilon_n)} x^2 \cdot \omega(x) \cdot p(x). \end{aligned} \quad (\text{B.3.2})$$

By Assumption 8 (i),(iii),

$$\lim_{n \rightarrow \infty} \sum_{x \in \mathcal{X} \setminus B_n} x^2 \cdot \omega(x) \cdot p(x) = 0 \quad (\text{B.3.3})$$

is true. Likewise, Assumption 8 (i),(iii) yield

$$\lim_{n \rightarrow \infty} \sum_{x \in B_n} x^2 \cdot \omega(x) \cdot p(x) = \sum_{x \in \mathcal{X}} x^2 \cdot \omega(x) \cdot p(x) =: c(\tau). \quad (\text{B.3.4})$$

Note that $c(\tau) \geq 0$, and it holds with equality only if $\omega(x) \cdot p(x) = 0$ for all $x \in \mathcal{X}$. But it is not true that $\omega(x) \cdot p(x) = 0$ for all $x \in \mathcal{X}$, from the property of (α, β) that it uniquely solves $\min_{(a,b)} E[\omega(X)(M(X) - a - bX)]$. Thus, it follows from (B.3.1)–(B.3.4) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha, \beta + \epsilon_n) \cdot \omega(X)]}{\epsilon_n^2} \\ \geq c(\tau) > 0 \end{aligned}$$

is true. But then, there exists $n^* > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha, \beta + \epsilon_{n^*}) \cdot \omega(X)]}{\epsilon_{n^*}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] > E[(\Delta^2(X, \alpha, \beta + \epsilon_{n^*}) \cdot \omega(X)],$$

and it contradicts Assumption 8 (ii), that (α, β) uniquely solves $\min_{(a,b)} E[\omega(X) \cdot (M(X) - a - bX)^2]$.

Next, we consider case (II). Suppose there exists x^{**} such that $\Delta(x^{**}, \alpha, \beta) = 0$. If $x^* = \max \mathcal{X}$, we can derive a contradiction from an argument similar to case (I). If $x^* \neq \max \mathcal{X}$, then since $\beta - \Delta M(x) < 0$ for all $x \in \mathcal{X}$, $\Delta(x, \alpha, \beta) < 0$ for all $x > x^{**}$. This contradicts $\Delta(x, \alpha, \beta) \geq 0$ for all $x \in \mathcal{X}$. Hence, $\Delta(x, \alpha, \beta) > 0$ for all $x \in \mathcal{X}$. For each $\epsilon > 0$, define the set

$$A'(\epsilon) = \{x \in \mathcal{X} \mid \Delta(x, \alpha - \epsilon, \beta) \geq 0\}.$$

Let B denote an arbitrary nonempty finite subset of \mathcal{X} . If we choose $\epsilon := \min_{x \in B} \{\Delta(x, \alpha, \beta)\}$, then $\Delta(x, \alpha - \epsilon, \beta) \geq 0$ for all $x \in B$. Hence, for any finite subset $B \subset \mathcal{X}$, there exists $\epsilon > 0$ such that $B \subset A'(\epsilon)$. Furthermore, note that $\epsilon > 0$ is true. It follows that the inequality

$$\begin{aligned} E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[\Delta^2(X, \alpha - \epsilon, \beta) \cdot \omega(X)] \\ = \sum_{x \in \mathcal{X}} (2\Delta(x, \alpha, \beta)\epsilon - \epsilon^2) \cdot \omega(x) \cdot p(x) \\ \geq \sum_{x \in A'(\epsilon)} \epsilon^2 \cdot \omega(x) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus A'(\epsilon)} \epsilon^2 \cdot \omega(x) \cdot p(x) \end{aligned} \quad (\text{B.3.5})$$

holds for any $\epsilon > 0$. Let $B_n = \{x \in \mathcal{X} \mid -n \leq x \leq n\}$ for each integer $n > 1$. Then, for each $n > 1$, there exists ϵ_n such that $B_n \subset A'(\epsilon_n)$. But then, for each $n > 1$, there exists $\epsilon_n > 0$ such

that,

$$\begin{aligned} & \sum_{x \in B_n} \omega(x) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus B_n} \omega(x) \cdot p(x) \\ & \leq \sum_{x \in A'(\epsilon_n)} \omega(x) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus A'(\epsilon_n)} \omega(x) \cdot p(x). \end{aligned} \quad (\text{B.3.6})$$

Since $\sum_{x \in \mathcal{X}} \omega(x) \cdot p(x) < \infty$,

$$\lim_{n \rightarrow \infty} \sum_{x \in \mathcal{X} \setminus B_n} \omega(x) \cdot p(x) = 0 \quad \text{and} \quad (\text{B.3.7})$$

$$\lim_{n \rightarrow \infty} \sum_{x \in B_n} \omega(x) \cdot p(x) = \sum_{x \in \mathcal{X}} \omega(x) \cdot p(x) = c'(\tau) > 0. \quad (\text{B.3.8})$$

Thus, it follows from (B.3.5)–(B.3.8) that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha - \epsilon_n, \beta) \cdot \omega(X)]}{\epsilon_n^2} \\ & \geq c'(\tau) > 0 \end{aligned}$$

is true. But then, there exists $n^{**} > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha - \epsilon_{n^{**}}, \beta) \cdot \omega(X)]}{\epsilon_{n^{**}}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] > E[(\Delta^2(X, \alpha - \epsilon_{n^{**}}, \beta) \cdot \omega(X)],$$

and it contradicts Assumption 8 (ii), that (α, β) uniquely solves $\min_{(a,b)} E[\omega(X) \cdot (M(X) - a - bX)^2]$. \square

B.4 Proof of Theorem 3

Proof. First, consider the case where there exists $\tilde{x} \in \mathcal{X}$ such that $\Delta L(\tilde{x}) - \Delta M(\tilde{x}) = 0$. In this case,

$$\beta = \Delta L(\tilde{x}) = \Delta M(\tilde{x}) = E[\tilde{w}(X) \cdot \Delta M(X)] \quad (\text{B.4.1})$$

holds, where $\tilde{w}(x) = 1/p(x)$ if $x = \tilde{x}$ and $\tilde{w}(x) = 0$ otherwise.

Next, consider the case where $\Delta L(x) - \Delta M(x) \neq 0$ for all $x \in \mathcal{X}$. By Lemma 2, there exist

$x_1, x_2 \in \mathcal{X}$ such that $(\Delta L(x_1) - \Delta M(x_1)) \cdot (\Delta L(x_2) - \Delta M(x_2)) < 0$. Thus,

$$\begin{aligned} \beta &= \frac{\beta - \Delta M(x_2)}{\Delta M(x_1) - \Delta M(x_2)} \cdot \Delta M(x_1) + \frac{\Delta M(x_1) - \beta}{\Delta M(x_1) - \Delta M(x_2)} \cdot \Delta M(x_2) \\ &= E[\lambda(X) \cdot \Delta M(X)] \end{aligned}$$

holds, where $\lambda(x_1) = \frac{\beta - \Delta M(x_2)}{\Delta M(x_1) - \Delta M(x_2)} \cdot \frac{1}{p(x_1)}$, $\lambda(x_2) = \frac{\Delta M(x_1) - \beta}{\Delta M(x_1) - \Delta M(x_2)} \cdot \frac{1}{p(x_2)}$, and $\lambda(x) = 0$ for all $x \in \mathcal{X} \setminus \{x_1, x_2\}$. Note that this weight function λ is non-negative because of $(\Delta L(x_1) - \Delta M(x_1)) \cdot (\Delta L(x_2) - \Delta M(x_2)) \leq 0$. \square

B.5 Proof of Lemma 3

Proof. Define the short-hand notation

$$\Delta(x, \alpha, \beta) := \alpha + x'\beta - M(x) = L(x) - M(x).$$

By way of contradiction, suppose $M \geq L$ for all $x \in \mathcal{X}$. Then, either one of the following two cases is true.

(I) There exists a point $x^* \in \mathcal{X}$ such that $\Delta(x^*, \alpha, \beta) = 0$.

(II) $\Delta(x, \alpha, \beta) < 0$ for all $x \in \mathcal{X}$.

We claim below that each of the cases (I)–(II) contradicts Assumption 9 that (α, β) uniquely solves $\min_{(a,b)} E[\omega(X) \cdot \Delta^2(X, \alpha, \beta)]$.

First, consider case (I). We show that x^* is unique. Suppose there exists $x^{**} \neq x^*$ such that $\Delta(x^{**}, \alpha, \beta) = 0$. Since M is strictly convex from Assumption 9 (vii), for any $0 < \lambda < 1$, $M(\lambda x^* + (1 - \lambda)x^{**}) < \lambda \cdot M(x^*) + (1 - \lambda)M(x^{**}) = \alpha + (\lambda x^* + (1 - \lambda)x^{**})'\beta$. This contradicts $M \geq L$ for all $x \in \mathcal{X}$. Hence, x^* is unique. Without loss of generality, we

normalize the location to $x^* = 0$. For each $\epsilon > 0$, define the set

$$A(\epsilon) = \{x \in \mathcal{X} \mid -\epsilon - \Delta(x, \alpha, \beta) \geq 0\}.$$

Let B^1 denote an arbitrary compact rectangle in \mathcal{X} and B^2 denote an open rectangle in \mathcal{X} such that $B^2 \subset B^1$ and $0 \in B^2$. Let $B := B^1 \setminus B^2$. Note that B is compact in \mathcal{X} and $\min_{x \in B} \{-\Delta(x, \alpha, \beta)\}$ exists due to the compactness of B and the continuity of $\Delta(x, \alpha, \beta)$ with respect to x on B .

If we choose $\epsilon := \min_{x \in B} \{-\Delta(x, \alpha, \beta)\}$, then $-\epsilon - \Delta(x, \alpha, \beta) \geq 0$ for all $x \in B$. Hence, for any B^1, B^2 which satisfies the aforementioned property, there exists $\epsilon > 0$ such that $B \subset A(\epsilon)$. Furthermore, note that $\epsilon > 0$ is true. Now observe from the definition of $A(\epsilon)$ that the inequality

$$\begin{aligned} E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha + \epsilon, \beta) \cdot \omega(X))] \\ = \int_{\mathcal{X}} (-2\Delta(x, \alpha, \beta)\epsilon - \epsilon^2) \cdot \omega(x) dF_X(x) \\ \geq \int_{A(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) \end{aligned} \quad (\text{B.5.1})$$

holds for any $\epsilon > 0$. Let $B_n^1 := [-n, n]^k$, $B_n^2 := (-\frac{1}{n}, \frac{1}{n})^k$, and $B_n = B_n^1 \setminus B_n^2$ for each integer $n > 1$. Then, for each integer $n > 1$, there exists $\epsilon_n > 0$ such that $B_n \subset A(\epsilon_n)$. But then, for each $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} \int_{B_n} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) \\ \leq \int_{A(\epsilon_n)} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon_n)} \omega(x) dF_X(x). \end{aligned} \quad (\text{B.5.2})$$

By Assumption 9 (iii) and the Monotone Convergence Theorem,

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) = 0 \quad \text{and} \quad (\text{B.5.3})$$

$$\lim_{n \rightarrow \infty} \int_{B_n} \omega(x) dF_X(x) = \int_{\mathcal{X}} \omega(x) \cdot f_X(x) dx =: c' > 0. \quad (\text{B.5.4})$$

Thus, it follows from (B.5.1)–(B.5.4) that

$$\lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[\Delta^2(X, \alpha + \epsilon_n, \beta) \cdot \omega(X)]}{\epsilon_n^2} \geq c' > 0$$

is true. But then, there exists $n^{**} > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[\Delta^2(X, \alpha + \epsilon_{n^{**}}, \beta) \cdot \omega(X)]}{\epsilon_{n^{**}}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] > E[\Delta^2(X, \alpha + \epsilon_{n^{**}}, \beta) \cdot \omega(X)],$$

and it contradicts Assumption 9 (ii), that β uniquely minimizes the weighted mean squared distance.

Next, we consider case (II). As in case (I), if B^1 and B^2 satisfy the property mentioned in case (I), then there exists $\epsilon > 0$ such that $(B^1 \setminus B^2) \subset A(\epsilon)$. Thus,

$$\begin{aligned} & E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[(\Delta^2(X, \alpha + \epsilon, \beta) \cdot \omega(X))] \\ & \geq \int_{A(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon)} \epsilon^2 \cdot \omega(x) dF_X(x) \end{aligned} \quad (\text{B.5.5})$$

holds for any $\epsilon > 0$ similarly to case (I). By the same argument as the one used in case (I), for each integer $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} & \int_{B_n} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) \\ & \leq \int_{A(\epsilon_n)} \omega(x) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon_n)} \omega(x) dF_X(x). \end{aligned} \quad (\text{B.5.6})$$

Assumption 9 (iii) and Monotone Convergence Theorem yield

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X} \setminus B_n} \omega(x) dF_X(x) = 0 \quad \text{and} \quad (\text{B.5.7})$$

$$\lim_{n \rightarrow \infty} \int_{B_n} \omega(x) dF_X(x) = \int_{\mathcal{X}} \omega(x) \cdot f_X(x) dx =: c'. \quad (\text{B.5.8})$$

Note that $c \geq 0$, and it holds with equality only if $\omega(x, \alpha, \beta\tau) \cdot f_X(x) = 0$ almost everywhere

on \mathcal{X} . But it is not true that $\omega(X) \cdot f_X(x) = 0$ almost everywhere on \mathcal{X} , from Assumption 9 (ii) that (α, β) uniquely minimizes the weighted mean squared distance.

Thus, it follows from (B.5.5)–(B.5.8) that

$$\lim_{n \rightarrow \infty} \frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[\Delta^2(X, \alpha + \epsilon_n, \beta) \cdot \omega(X)]}{\epsilon_n^2} \geq c' > 0$$

is true, similarly to case (II). But then, there exists $n^{**} > 1$ such that

$$\frac{E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] - E[\Delta^2(X, \alpha + \epsilon_{n^{**}}, \beta) \cdot \omega(X)]}{\epsilon_{n^{**}}^2} > 0.$$

This inequality implies that

$$E[\Delta^2(X, \alpha, \beta) \cdot \omega(X)] > E[\Delta^2(X, \alpha + \epsilon_{n^{**}}, \beta) \cdot \omega(X)],$$

and it contradicts Assumption 9 (ii), that β uniquely minimizes the weighted mean squared distance. \square

B.6 Proof of Lemma 4

Proof. From Assumption 9 (vii), there exist $x'_1 > x_1^*$ and $x''_1 < x_1^*$ such that $\frac{\partial M(x'_1, x_{-1}^*)}{\partial x_1} > \beta_1$ and $\frac{\partial M(x''_1, x_{-1}^*)}{\partial x_1} < \beta_1$.

If $M(x'_1, x_{-1}^*) > L(x'_1, x_{-1}^*)$, then by the Intermediate Value Theorem there exists $\tilde{x}_1 \in [x_1^*, x'_1]$ such that $M(\tilde{x}_1, x_{-1}^*) = L(\tilde{x}_1, x_{-1}^*)$. Now consider the case where $M(x'_1, x_{-1}^*) < L(x'_1, x_{-1}^*)$. Let $\bar{\beta}_1 := \frac{\partial M(x'_1, x_{-1}^*)}{\partial x_1}$. Recall that $\bar{\beta} > \beta$. From Assumption 9(vii), for all $x_1 > x'_1$,

$$M(x_1, x_{-1}^*) > \bar{\beta}_1(x_1 - x'_1) + x_{-1}^* \beta_{-1} + M(x'_1, x_{-1}^*). \quad (\text{B.6.1})$$

Note that $\bar{\beta}_1(x_1 - x'_1) + x_{-1}^* \beta_{-1} + M(x'_1, x_{-1}^*)$, as a function of x_1 , is the tangent line to the curve $M(\cdot, x_{-1}^*)$ at the point $x_1 = x'_1$. Since $\bar{\beta}_1 > \beta_1$, $\bar{\beta}_1(x_1 - x'_1) + x_{-1}^* \beta_{-1} + M(x'_1, x_{-1}^*)$ and $L(x_1, x_{-1}^*)$ cross at the point $x_1 = \frac{M(x'_1, x_{-1}^*)}{\bar{\beta} - \beta}$. Further note that $\frac{M(x'_1, x_{-1}^*)}{\bar{\beta} - \beta} > x'_1$. Hence, from equation (B.6.1), we have

$$M\left(\frac{M(x'_1, x_{-1}^*)}{\bar{\beta} - \beta}, x_{-1}^*\right) > \bar{\beta}_1 \left(\frac{M(x'_1, x_{-1}^*)}{\bar{\beta} - \beta} - x'_1\right) + x_{-1}^* \beta_{-1} + M(x'_1, x_{-1}^*) = L\left(\frac{M(x'_1, x_{-1}^*)}{\bar{\beta} - \beta}, x_{-1}^*\right).$$

From the Intermediate Value Theorem, there exists $\bar{x}_1 \in \left[x'_1, \frac{M(x'_1, x_{-1}^*)}{\bar{\beta} - \beta} \right]$ such that $M(\bar{x}_1, x_{-1}^*) = L(\bar{x}_1, x_1^*)$.

Similarly we can show that there exists $\underline{x}_1 < x_1^*$ such that $M(\underline{x}_1, x_{-1}^*) = L(\underline{x}_1, x_1^*)$. \square

B.7 Proof of Theorem 4

Proof. From Lemma 3, there exists $x^* \in \mathcal{X}$ such that $L(x^*) > M(x^*)$. Since M is continuous, there exists $\epsilon > 0$, such that for all $x \in B_\epsilon(x^*)$, $L(x) > M(x)$ where B_ϵ denotes the open ϵ -ball of $x^* \in \mathcal{X}$. Hence, $\int_{x \in \mathcal{D}} dx \geq \int_{x \in B_\epsilon} dx > 0$. Further note that $\int_{x_{-1} \in \mathcal{D}_{-1}} dx_{-1} > 0$.

From Lemma 3 and 4, for each $x_{-1} \in \mathcal{D}_{-1}$, there exist $x'_1(x_{-1}) \in \mathbb{R}$ and $x''_1(x_{-1}) \in \mathbb{R}$ such that $x'_1(x_{-1}) < x''_1(x_{-1})$, $L(x'_1(x_{-1}), x_{-1}) = M(x'_1(x_{-1}), x_{-1})$ and $L(x''_1(x_{-1}), x_{-1}) = M(x''_1(x_{-1}), x_{-1})$. Hence, from the Fundamental Theorem of Calculus, for all $x_{-1} \in \mathcal{D}_{-1}$,

$$\int_{x'_1(x_{-1})}^{x''_1(x_{-1})} \frac{\partial M(x_1(x_{-1}), x_{-1})}{\partial x_1} dx_1 = \beta \cdot (x''_1(x_{-1}) - x'_1(x_{-1})).$$

By Assumption 9 (v), $f_X > 0$ almost everywhere on \mathcal{D} . Also recall that $\int_{x_{-1} \in \mathcal{D}_{-1}} dx_{-1} > 0$.

Thus, we can write

$$\begin{aligned} \beta_1 &= \frac{\int_{x_{-1} \in \mathcal{D}_{-1}} \int_{x'_1(x_{-1})}^{x''_1(x_{-1})} \frac{1}{(x''_1(x_{-1}) - x'_1(x_{-1}))} \cdot \frac{\partial M(x_1, x_{-1})}{\partial x_1} dx_1 dx_{-1}}{\int_{x_{-1} \in \mathcal{D}_{-1}} dx_{-1}} \\ &= \frac{\int_{x \in \mathcal{D}} \frac{1}{(x''_1(x_{-1}) - x'_1(x_{-1}))} \cdot \frac{\partial M(x)}{\partial x_1} \cdot \frac{1}{f_X(x)} dF_X(x)}{\int_{x_{-1} \in \mathcal{D}_{-1}} \frac{1}{f_{X_{-1}}} dF_{X_{-1}}(x_{-1})} \\ &= E \left[w_{\mathcal{D}}(X) \cdot \frac{\partial M(X)}{\partial X_1} \middle| X \in \mathcal{D} \right]. \end{aligned}$$

\square

B.8 Proof of Lemma 6

Proof. Define the function $P_\tau : \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{X}) \rightarrow \mathbb{R}$ by

$$P_\tau(\Gamma) = \int \mu_\tau(\omega(\Gamma, x), x) dF_X(x)$$

for each $\Gamma \in \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{X})$, where $\omega(\Gamma, x)$ is defined by

$$\omega(\Gamma, x) = \{u \in \mathcal{U} \mid (u, x) \in \Gamma\}.$$

Note that this set $\omega(\Gamma, x)$ belongs to $\mathcal{B}(\mathcal{U})$, as Γ belongs to the tensor product sigma algebra $\mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{X})$. We first show that this function P_τ is a probability measure. That $P_\tau(\Gamma) \in [0, 1]$ for all $\Gamma \in \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{X})$ follows from the fact that $\mu_{y,x}$ is a probability measure (Lemma 5) and F_X is a probability measure. Likewise, by Lemma 5, we have

$$P_\tau(\emptyset) = \int \mu_{Q_\tau(x),x}(\emptyset) dF_X(x) = 0 \quad \text{and}$$

$$P_\tau(\mathcal{U} \times \mathcal{X}) = \int \mu_{Q_\tau(x),x}(\partial V(Q_\tau(x), x)) dF_X(x) = \int_{\mathcal{X}} dF_X(x) = 1.$$

Let $\{\Gamma_i\}_i \subset \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{X})$ be a countable collection of disjoint sets. Note that, for each $x \in \mathcal{X}$, $\omega(\Gamma_i, x) \cap \omega(\Gamma_j, x) = \emptyset$ whenever $i \neq j$, and $\omega(\cup_i \Gamma_i, x) = \cup_i \omega(\Gamma_i, x)$. Therefore, from the fact that $\mu_{y,x}$ is a probability measure (Lemma 5), we have the sigma additivity

$$\begin{aligned} P_\tau(\cup_i \Gamma_i) &= \int \mu_\tau(\omega(\cup_i \Gamma_i, x), x) dF_X(x) = \int \mu_\tau(\cup_i \omega(\Gamma_i, x), x) dF_X(x) \\ &= \int \sum_i \mu_\tau(\omega(\Gamma_i, x), x) dF_X(x) = \sum_i P_\tau(\Gamma_i) \end{aligned}$$

where the last step uses the Fubini-Tonelli Theorem.

To show that μ_τ is a regular conditional probability measure given the events in $\mathcal{B}(\mathcal{X})$, it

remains to show that

$$P_\tau(S \times T) = \int_T \mu_\tau(S, x) dF_X(x)$$

holds for all $(S, T) \in \mathcal{B}(\mathcal{U}) \times \mathcal{B}(\mathcal{X})$. This follows straightforwardly, as

$$\begin{aligned} P_\tau(S \times T) &= \int \mu_\tau(\omega(S \times T, x), x) dF_X(x) \\ &= \int \mu_\tau(\{u \in \mathcal{U} \mid (u, x) \in S \times T\}, x) dF_X(x) = \int_T \mu_\tau(S, x) dF_X(x). \end{aligned}$$

□

Appendix C

Mathematical Appendix of Chapter 3

C.1 Proof of Theorem 7

Proof. We show that $(\alpha(\tau), \beta(\tau))$ which solve the problem

$$\min_{\beta} (E[(1\{Y < \alpha(\tau) + \beta X\} - \tau) \cdot Z])^2 \quad (\text{C.1.1})$$

is equivalent to the $(\alpha(\tau), \beta(\tau))$ which solves the following problem.

$$\min_{\beta} E[\tilde{\omega}_{\tau}(X, \alpha(\tau), \beta(\tau)) \cdot \Delta_{\tau}^2(X, \alpha(\tau), \beta)] \quad (\text{C.1.2})$$

The first order condition of the equation (C.1.1) is

$$(E[(1\{Y < \alpha(\tau) + \beta X\} - \tau) \cdot Z]) = 0$$

$$\begin{aligned}
E[(1\{Y < \alpha(\tau) + \beta X\} - \tau) \cdot Z] &= E[(1\{\epsilon_\tau < \Delta_\tau(X, \alpha(\tau), \beta)\} - \tau) \cdot Z] \\
&= E[(F_{\epsilon_\tau|XZ}(\Delta_\tau(X, \alpha(\tau), \beta)|XZ) - F_{\epsilon_\tau|XZ}(0|XZ)) \cdot Z] \\
&= E[(\int_0^1 f_Y(u\Delta_\tau(X, \alpha(\tau), \beta)|XZ)du) \cdot \Delta_\tau(X, \alpha(\tau), \beta) \cdot \frac{Z}{X} \cdot X] \\
&= E[\tilde{\omega}_\tau(X, Z, \alpha(\tau), \beta(\tau)) \cdot \Delta_\tau(X, \alpha(\tau), \beta) \cdot X] \\
&= 0
\end{aligned} \tag{C.1.3}$$

Which is equivalent to the first order condition of (C.1.2). Hence $\beta(\tau)$ is the solution to both equation (C.1.2) and equation (C.1.1).

□

C.2 Proof of Lemma 7

Proof. Since $Q_\tau(0)$ is known, $\alpha(\tau) = Q_\tau(0)$. Hence L_τ and Q_τ intersect at $x=0$. Suppose L_τ and Q_τ only intersect at the point $x=0$. Without loss of generality, assume $L_\tau > Q_\tau$ for all $x \in \mathcal{X} \setminus \{0\}$. Let $\tilde{\omega}(x, \alpha(\tau), \beta(\tau)) = \int_{\mathcal{Z}} \tilde{\omega}(x, z, \alpha(\tau), \beta(\tau)) dF_Z(z|x)$. Then the slope of the linear quantile regression $\beta(\tau)$ solves

$$\min_{\beta} E[\tilde{\omega}_\tau(X, \alpha(\tau), \beta(\tau)) \cdot \Delta_\tau^2(X, \alpha(\tau), \beta)]$$

From Assumption 13 (ii) and (v), $\tilde{\omega}(x, \alpha(\tau), \beta(\tau)) \geq 0$ for all $x \in \mathcal{X}$.

For each $\epsilon > 0$, define the set

$$A(\epsilon) = \{x \in \mathcal{X} \setminus \{0\} \mid \Delta_\tau(x, \alpha(\tau), \beta(\tau) + \epsilon) \leq 0\}$$

We let B denote arbitrary compact intervals contained in $\mathcal{X} \setminus \{0\}$. Note that $\min_{x \in B} \left\{ \frac{\Delta_\tau(x, \alpha(\tau), \beta(\tau))}{-x} \right\}$

exists due to the compactness of B and the continuity of $\frac{\Delta_\tau(x, \alpha(\tau), \beta(\tau))}{-x}$ with respect to x on B . Thus, if we choose $\epsilon := \min_{x \in B} \left\{ \frac{\Delta_\tau(x, \alpha(\tau), \beta(\tau))}{-x} \right\}$, then $\Delta_\tau(x, \alpha(\tau), \beta(\tau) + \epsilon) \geq 0$ for all $x \in B$. Hence, for any compact intervals $B \subset \mathcal{X} \setminus \{0\}$ there exists $\epsilon > 0$ such that $B \subset A(\epsilon)$. Furthermore, note that $\epsilon > 0$ is true. Now, observe from the definition of $A(\epsilon)$ that the inequality

$$\begin{aligned} & E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] - E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)))] \\ &= \int_{\mathcal{X}} (-2\Delta_\tau(x, \alpha(\tau), \beta(\tau))\epsilon x - (\epsilon x)^2) \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) \\ &\geq \int_{A(\epsilon)} (\epsilon x)^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon)} (\epsilon x)^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) \end{aligned} \quad (\text{C.2.1})$$

holds for any $\epsilon > 0$. Let $B_n := \mathcal{X} \cap [\frac{1}{n}, n]$ for each integer $n > 1$. Then for each $n > 1$, the above argument implies that there exists $\epsilon_n > 0$ such that $B_n \subset A(\epsilon_n)$. But then, for each $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} & \int_{B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) - \int_{\mathcal{X} \setminus B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) \\ &\leq \int_{A(\epsilon_n)} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) - \int_{\mathcal{X} \setminus A(\epsilon_n)} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x). \end{aligned} \quad (\text{C.2.2})$$

By Assumption 12 (i)–(ii) and the monotone convergence theorem,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int_{\mathcal{X} \setminus B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) \\ &= \lim_{n \rightarrow \infty} \int_{\mathcal{X}} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot f_X(x) \cdot \mathbb{1}\{x \in \mathcal{X} \setminus B_n\} dx \\ &= \int_{\mathcal{X}} \lim_{n \rightarrow \infty} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot f_X(x) \cdot \mathbb{1}\{x \in \mathcal{X} \setminus B_n\} dx = 0 \end{aligned} \quad (\text{C.2.3})$$

is true. Likewise, Assumption 12 (i)–(ii) and the monotone convergence theorem yield

$$\lim_{n \rightarrow \infty} \int_{B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) dF_X(x) = \int_{\mathcal{X}} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot f_X(x) dx =: c(\tau). \quad (\text{C.2.4})$$

Note that $c(\tau) \geq 0$ and, it holds with equality only if $\bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot f_X(x) = 0$ almost everywhere on \mathcal{X} . But it is not true that $\bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot f_X(x) = 0$ almost everywhere on \mathcal{X} from the property of $(\beta(\tau))$ that it uniquely solves (C.1.1). Thus it follows from (C.2.1)–(C.2.4) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] - E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon_n) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)))]}{\epsilon_n^2} \\ \geq c(\tau) > 0 \end{aligned}$$

is true. But then, there exists $n^* > 1$ such that

$$\frac{E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] - E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon_{n^*}) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)))]}{\epsilon_{n^*}^2} > 0.$$

This inequality implies

$$E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] > E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon_{n^*}) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))],$$

and it contradicts the aforementioned property of $\beta(\tau)$ that it solves (C.1.2). □

C.3 Proof of Theorem 8

Proof. From Lemma 7, Assumption 12 guarantees existence of two points $0, x_1 \in \mathcal{X}$ such that

$L_\tau(0) = Q_\tau(0)$ and $L_\tau(x_1) = Q_\tau(x_1)$ are both true. By Assumption 3, we have $Q_\tau(x_1) -$

$Q_\tau(0) = \int_0^{x_1} Q'_\tau(\xi) d\xi$ by the Fundamental Theorem of Calculus. Similarly, we have $L_\tau(x_1) -$

$L_\tau(0) = \int_0^{x_1} L'_\tau(\xi) d\xi = \beta(\tau) \cdot (x_1 - 0)$. Combining all these equalities together yields

$$\int_0^{x_1} Q'_\tau(\xi) d\xi = \beta(\tau) \cdot x_1.$$

By Assumption 12 (v), $f_X > 0$ almost everywhere on $[0, x_1]$. Thus, we can write

$$\beta(\tau) = \frac{\int_0^{x_1} Q'_\tau(\xi) d\xi}{\int_0^{x_1} d\xi} = \frac{\int_0^{x_1} \frac{Q'_\tau(\xi)}{f_X(\xi)} dF_X(\xi)}{\int_0^{x_1} \frac{1}{f_X(\xi)} dF_X(\xi)} = E[w_{x_1}(X) \cdot Q'_\tau(X) | 0 \leq X \leq x_1].$$

□

C.4 Proof of Lemma 8

Proof. By way of contradiction, suppose that $\Delta L_\tau(x) - \Delta Q_\tau(x) < 0$ is true for all $x \in \mathcal{X}$. From Theorem 7, if Assumption 15 (i)–(v) are true, then $\beta(\tau)$ uniquely solves the weighted least squares problem

$$\min_{\beta} E[\tilde{\omega}_\tau(X, Z, \alpha(\tau), \beta(\tau)) \cdot \Delta_\tau^2(X, \alpha(\tau), \beta)], \quad (\text{C.4.1})$$

where $\tilde{\omega}_\tau(x, z, \alpha(\tau), \beta(\tau))$ and $\Delta_\tau(X, \alpha, \beta)$ are defined by

$$\tilde{\omega}_\tau(x, z, \alpha(\tau), \beta(\tau)) := \frac{Z}{X} \cdot \int_0^1 f_Y(u(\alpha(\tau) + \beta(\tau)x) + (1-u)Q_{Y|X}(\tau|x)|XZ)du,$$

$$\Delta_\tau(X, \alpha, \beta) := \alpha(\tau) + \beta(\tau)x - Q_\tau(x) = L_\tau(x) - Q_\tau(x).$$

Let $\bar{\omega}(x, \alpha(\tau), \beta(\tau)) = \int_{\mathcal{Z}} \tilde{\omega}(x, z, \alpha(\tau), \beta(\tau)) dF_Z(z|x)$. Then the slope of the linear quantile regression $\beta(\tau)$ solves

$$\min_{\beta} E[\bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)) \cdot \Delta_\tau^2(X, \alpha(\tau), \beta)]$$

From Assumption 15 (ii) and (v), $\bar{\omega}(x, \alpha(\tau), \beta(\tau)) \geq 0$ for all $x \in \mathcal{X}$.

Since $\Delta L_\tau(x) - \Delta Q_\tau(x) < 0$ for all $x \in \mathcal{X}$, if $\Delta_\tau(x, \alpha(\tau), \beta(\tau)) < 0$ for some $x \in \mathcal{X}$, then $\Delta_\tau(\tilde{x}, \alpha(\tau), \beta(\tau)) < 0$ for all $\tilde{x} > x$. Also, if $\Delta_\tau(x', \alpha(\tau), \beta(\tau)) > 0$ for some $x' \in \mathcal{X}$, then $\Delta_\tau(\tilde{x}, \alpha(\tau), \beta(\tau)) > 0$ for all $\tilde{x} < x'$. From Assumption 15 (i) $L(0) - Q(0) = 0$. Hence $\Delta_\tau(x, \alpha(\tau), \beta(\tau)) \geq 0$ for all $x \leq 0$ and $\Delta_\tau(x, \alpha(\tau), \beta(\tau)) < 0$ for all $x > 0$. For each $\epsilon > 0$, define the set

$$A(\epsilon) = \{x \in \mathcal{X} \cap \mathbb{Z}_+ \mid \Delta_\tau(x, \alpha(\tau), \beta(\tau) + \epsilon) \leq 0\} \cup \{x \in \mathcal{X} \cap \mathbb{Z}_- \mid \Delta_\tau(x, \alpha(\tau), \beta(\tau) + \epsilon) \geq 0\}.$$

We let B denote an arbitrary nonempty finite subset of \mathcal{X} .

If we choose $\epsilon := \min_{x \in B} \left\{ \frac{\Delta_\tau(x, \alpha(\tau), \beta(\tau))}{-x} \right\}$, then $\Delta_\tau(x, \alpha(\tau), \beta(\tau) + \epsilon) \geq 0$ for all $x \in$

$B \cap \mathcal{X}_+$ and $\Delta_\tau(x, \alpha(\tau), \beta(\tau) + \epsilon) \leq 0$ for all $x \in B \cap \mathcal{X}_-$. Hence, for any finite subset B , there exists $\epsilon > 0$ such that $B \subset A(\epsilon)$. Furthermore, note that $\epsilon > 0$ is true. Now, observe from the definition of $A(\epsilon)$ that the inequality

$$\begin{aligned} & E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] - E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)))] \\ &= \sum_{x \in \mathcal{X}} (-2\Delta_\tau(x, \alpha(\tau), \beta(\tau))\epsilon x - (\epsilon x)^2) \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) \\ &\geq \sum_{x \in A(\epsilon)} (\epsilon x)^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus A(\epsilon)} (\epsilon x)^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) \quad (\text{C.4.2}) \end{aligned}$$

holds for any $\epsilon > 0$. Let $B_n := \{x \in \mathcal{X} \mid -n \leq x \leq n\}$ for each integer $n > 1$. Then for each $n > 1$, the above argument implies that there exists $\epsilon_n > 0$ such that $B_n \subset A(\epsilon_n)$. But then, for each $n > 1$, there exists $\epsilon_n > 0$ such that

$$\begin{aligned} & \sum_{x \in B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) \\ &\leq \sum_{x \in A(\epsilon_n)} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) - \sum_{x \in \mathcal{X} \setminus A(\epsilon_n)} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x). \quad (\text{C.4.3}) \end{aligned}$$

By Assumption 15 (i)–(ii),

$$\lim_{n \rightarrow \infty} \sum_{x \in \mathcal{X} \setminus B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) = 0 \quad (\text{C.4.4})$$

is true. Likewise, Assumption 15 (i)–(ii) yield

$$\lim_{n \rightarrow \infty} \sum_{x \in B_n} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) = \sum_{x \in \mathcal{X}} x^2 \cdot \bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) =: c(\tau). \quad (\text{C.4.5})$$

Note that $c(\tau) \geq 0$ and, it holds with equality only if $\bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) = 0$ for all $x \in \mathcal{X}$. But it is not true that $\bar{\omega}_\tau(x, \alpha(\tau), \beta(\tau)) \cdot p(x) = 0$ for all $x \in \mathcal{X}$ from the property of $(\alpha(\tau), \beta(\tau))$ that it uniquely solves (C.4.1). Thus it follows from (C.4.2)–(C.4.5) that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] - E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon_n) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)))]}{\epsilon_n^2} \\ &\geq c(\tau) > 0 \end{aligned}$$

is true. But then, there exists $n^* > 1$ such that

$$\frac{E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] - E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon_{n^*}) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau)))]}{\epsilon_{n^*}^2} > 0.$$

This inequality implies

$$E[\Delta_\tau^2(X, \alpha(\tau), \beta(\tau)) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))] > E[(\Delta_\tau^2(X, \alpha(\tau), \beta(\tau) + \epsilon_{n^*}) \cdot \bar{\omega}_\tau(X, \alpha(\tau), \beta(\tau))],$$

and it contradicts the aforementioned property of $\beta(\tau)$ that it solves (C.4.1). \square

C.5 Proof of Theorem 9

Proof. First, consider the case where there exists $\tilde{x} \in \mathcal{X}$ such that $\Delta L_\tau(\tilde{x}) - \Delta Q_\tau(\tilde{x}) = 0$. In this case,

$$\beta(\tau) = \Delta L_\tau(\tilde{x}) = \Delta Q_\tau(\tilde{x}) = E[w(X) \cdot \Delta Q_\tau(X)] \quad (\text{C.5.1})$$

holds, where $w(x) = 1/p(x)$ if $x = \tilde{x}$ and $w(x) = 0$ otherwise.

Next, consider the case where $\Delta L_\tau(x) - \Delta Q_\tau(x) \neq 0$ for all $x \in \mathcal{X}$. By Lemma 2, there exist $x_1, x_2 \in \mathcal{X}$ such that $(\Delta L_\tau(x_1) - \Delta Q_\tau(x_1)) \cdot (\Delta L_\tau(x_2) - \Delta Q_\tau(x_2)) < 0$. Thus,

$$\begin{aligned} \beta(\tau) &= \frac{\beta(\tau) - \Delta Q_\tau(x_2)}{\Delta Q_\tau(x_1) - \Delta Q_\tau(x_2)} \cdot \Delta Q_\tau(x_1) + \frac{\Delta Q_\tau(x_1) - \beta(\tau)}{\Delta Q_\tau(x_1) - \Delta Q_\tau(x_2)} \cdot \Delta Q_\tau(x_2) \\ &= E[\lambda(X) \cdot \Delta Q_\tau(X)] \end{aligned}$$

holds, where $\lambda(x_1) = \frac{\beta(\tau) - \Delta Q_\tau(x_2)}{\Delta Q_\tau(x_1) - \Delta Q_\tau(x_2)} \cdot \frac{1}{p(x_1)}$, $\lambda(x_2) = \frac{\Delta Q_\tau(x_1) - \beta(\tau)}{\Delta Q_\tau(x_1) - \Delta Q_\tau(x_2)} \cdot \frac{1}{p(x_2)}$, and $\lambda(x) = 0$ for all $x \in \mathcal{X} \setminus \{x_1, x_2\}$. Note that this weight function λ is non-negative because of $(\Delta L_\tau(x_1) - \Delta Q_\tau(x_1)) \cdot (\Delta L_\tau(x_2) - \Delta Q_\tau(x_2)) \leq 0$. \square

Ryutah Kato

Johns Hopkins University
Department of Economics
Wyman Park Building 544E
3400 N. Charles St.
Baltimore, MD 21218, USA

Phone: (443) 683-5852
Email: rkato3@jhu.edu
Citizenship: Japan

Education

B.S. Geophysics, Kyoto University, 2010.
M.A. Economics, University of Tokyo, 2013.
M.A. Economics, Johns Hopkins University, 2015.
Ph.D. Economics, Johns Hopkins University, 2020 (expected).

Research Interest

Econometrics, Corporate Finance, Industrial Organization

Academic Experience

Teaching Assistant

Undergraduate level: International Trade, Labor Economics, Corporate Finance
Graduate level: Microeconomic Theory, Statistical Inference

Publications

"On Using Linear Quantile Regressions for Causal Inference." with Yuya Sasaki, *Econometric Theory* 33.3 (2017): 664-690.

Skills

Programming: R, Matlab
Languages: Japanese (native), English (fluent)